Risk prediction in rheumatoid arthritis

Scott, Ian Clifford

Awarding institution:
King's College London

RISK PREDICTION IN RHEUMATOID ARTHRITIS

Ian Clifford Scott

Thesis submitted for the degree of Doctor of Philosophy

King's College London

2014

ABSTRACT

As rheumatoid arthritis (RA) is a heterogeneous disease whose course and treatment response varies between patients, a stratified approach to its management is required. This thesis aimed to facilitate the risk prediction that underpins stratified medicine in RA. Its primary aim was to improve the knowledge of which clinical and genetic factors predict RA's onset, disease course and treatment responses. Its secondary aim was to develop a prediction modelling framework that harnessed these factors to inform clinical care. There were five key findings.

Firstly, it demonstrated a significant inverse association between alcohol consumption and RA development when the evidence across published studies was pooled using meta-analytical techniques. This suggests alcohol may protect against RA. Secondly, it demonstrated that only HLA RA susceptibility variants associated with radiological progression in a clinical trial cohort of early, active RA patients. This suggests the non-HLA genetic architectures of RA susceptibility and severity may, at least partially, differ. Thirdly, it provided evidence that anti-citrullinated protein antibodies (ACPA) can identify patients with early, active RA that are most likely to benefit from combination treatments. Fourthly, it demonstrated that estimating an asymptomatic individual's risk of RA is possible, through developing and validating a risk prediction model that uses computer simulation to improve upon the discriminative abilities of existing RA prediction models. Finally, it highlighted the importance of considering RA's heterogeneity when assessing its predictive factors; alcohol's likely protective effect was predominantly seen in ACPA-positive disease and genetic and environmental factors had different impacts on the risk of developing younger and older onset RA.

In conclusion this thesis has contributed to stratified medicine in RA by better characterising which predictive factors are relevant to its development, severity, treatment needs and responses and developing a risk prediction modelling framework that may be applicable to many aspects of stratified care.

TABLE OF CONTENTS

SUMMARY OF TABLES

SUMMARY OF FIGURES

LIST OF ABBREVIATIONS

ACPA = Anti-citrullinated protein antibodies

ACR = American College of Rheumatology

AHRQ = Agency for Health Care Research and Quality

AIC = Akaike Information Criterion

Anti-TNF= Anti-tumour necrosis factor

APC = Antigen presenting cell

APIPPRA = Arthritis Prevention In the Pre-clinical Phase of RA with Abatacept

ARA = American Rheumatism Association

AS = Ankylosing spondylitis

AUC = Area under the curve

BAL = Bronchoalveolar lavage

BMI = Body mass index

BRASS = Brigham RA Sequential Study

BSRBR = British Society for Rheumatology Biologics Register

CACORA = Case-Control study in Rheumatoid Arthritis

CARDERA = Combination Anti-Rheumatic Drugs in Early RA

CGC = CARDERA Genetics Cohort

CI = Confidence interval

CLEAR = Consortium for the Longitudinal Evaluation of African Americans with Early RA

COMP = Cartilage oligomeric matrix protein

CRP = C-reactive protein.

CTX-II = C-telopeptide of type II collagen

DAS28 = Disease activity score on a 28-joint count

DMARD = Disease-modifying anti-rheumatic drug

DMP = Differentially methylated position

EAC = Early Arthritis Clinic

EIRA = Epidemiological Investigation of Rheumatoid Arthritis

EMS = Early morning stiffness

EORA = Elderly onset RA

ERAS = Early Rheumatoid Arthritis Study

ESR = Erythrocyte sedimentation rate

EULAR = European League Against Rheumatism

EUR = European

FDR = First-degree relative

FLS = Fibroblast-like synoviocytes

GENRA = GENetics of Ra in individuals of African ancestry

GORA = Genetics of RA

GRCh37 = Genome Reference Consortium Human Build 37

GWAS = Genome-wide association study

HAQ = Health assessment questionnaire

His13 = Histidine at position 13

HR = Hazard ratio

HRCT = High resolution computed tomography

HRT = Hormone replacement therapy

HWE = Hardy-Weinberg Equilibrium

IBD = Inflammatory bowel disease

IL1B = Interleukin 1, beta

IL1RN = Interleukin 1 receptor antagonist

IM = Intramuscular

INT = Inverse normal transformation

IP = Inflammatory polyarthritis

IRR = Incidence rate ratio

JAK = Janus-associated kinase

LASSO = Least absolute shrinkage and selection operator

LD = Linkage disequilibrium

MAF = Minor allele frequency

MAPK = Mitogen-activated protein kinase

MAR = Missing-at-random

MCID = Minimal clinically important difference

MCP = Metacarpophalangeal

MCS = Mental component summary

MHC = Major histocompatibility complex

MOOSE = Meta-analysis Of Observational Studies in Epidemiology

MRI = Magnetic resonance imaging

MTP = Metatarsophalangeal

NARAC = North American Rheumatoid Arthritis Consortium

NDB = National Databank for Rheumatic diseases

NF-kB = Nuclear factor-kB

NHS = Nurses' Health Study

NICE = National Institute for Health and Care Excellence

NOAR = Norfolk Arthritis Register

NRI = Net reclassification improvement

OA = Osteoarthritis

OCP = Oral contraceptive pill

OR = Odds ratio

P4 = Pocket 4

PAD = Peptidylarginine deiminase

PC = Principal component

PCA = Principal component analysis

PCRSSOP = Polymerase chain reaction-sequence-specific oligonucleotide probe

PCS = Physical component summary

PD = Periodontitis

PIAS = Protein inhibitor of activated STAT

PIP = Proximal interphalangeal

PPAD = Porphyromonas gingivalis peptidylarginine deiminase

PREVeNT RA = PRe-clinical EValuation of Novel Targets in RA

PTK = Protein tyrosine kinase

QC = Quality control

QoL = Quality of life

QQ = Quantile-quantile

RA = Rheumatoid Arthritis

RCT = Randomised controlled trial

REGENT = Risk Estimation for Genetic and Environmental Traits

RF = Rheumatoid factor

ROC = Receiver operating characteristic

RR = Relative risk

RRP = Rapid radiological progression

SE = Shared epitope

SJC = Swollen joint count

SNP = Single nucleotide polymorphism

SPAG16 = Sperm-Associated AntiGen 16

SvHS = Sharp/van der Heijde Score

TIMP-1 = Tissue inhibitor of metalloproteinases-1

TJC = Tender joint count

TNF = Tumour necrosis factor

UKRAGG = UK RA Genetics Group

USS = Ultrasound scanning

Val11 = Valine at position 11

VAS = Visual analogue scale

VEIA = Very early undifferentiated inflammatory polyarthritis

wGRS = Weighted genetic risk score

WTCCC = Wellcome Trust Case Control Consortium

YEAR = Yorkshire Early Arthritis Register

YORA = Younger onset RA

ZMIZ1 = Zinc finger, MIZ-type containing 1

ΔLarsen = 24-month change in Larsen score

λ = Genomic inflation factor

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my PhD supervisors, Professor Cathryn Lewis, Professor Andrew Cope and Dr Sophia Steer for their help, advice and support throughout the course of my PhD.

I would also like to acknowledge and thank the following individuals for their input into the research reported in this thesis. Chapter 2 was undertaken with statistical guidance from Dr Daniel Stahl, whose expertise in the meta-analysis of observational studies was crucial to performing this work. Chapters 3 and 4 were undertaken with input from Dr Sarah Spain (who undertook the genotype calling and principal components analysis), Mr Jelmar Quist (who undertook the HLA imputation) and Dr Jemma Walker, who provided advice on the linear mixed-effects modelling used. Chapter 5 was undertaken with input from Mr Seth Seegobin (who performed the repeated measures ANOVA modelling alongside me), Dr Margaret Ma and Mr Chanaka Dahanayake (who both assisted with the serological tests). Chapter 6 was undertaken as a collaborative project, with members of the UK RA Genetics Group and Welcome Trust Case Control Consortium contributing their data; without this, the study would not have been possible. Additionally the age of onset analysis reported in chapter 6 was undertaken by Mr Seth Seegobin.

I would like to express my appreciation to Arthritis Research UK for funding this Clinical Research Fellowship and enabling me to present these data at both national and international meetings.

Finally, I would like to thank my wife, Charlotte and daughter, Emma for their support and encouragement throughout my PhD.

PUBLICATIONS

Original Research Papers

Seegobin SD, Ma MHY, Dahanayake C, Cope AP, Scott DL, Lewis CM, **Scott IC**. ACPA-Positive And ACPA-Negative Rheumatoid Arthritis Differ In Their Requirements For Combination DMARDs and Corticosteroids: Secondary Analysis Of A Randomised Controlled Trial. Arthritis Res Ther 2014; 16: R13.

Scott IC, Seegobin SD, Steer S, Tan R, Forabosco P, Hinks A, Eyre S, Morgan AW, Wilson AG, Hocking LJ, Wordsworth P, Barton A, Worthington J, Cope AP, Lewis CM. Predicting the Risk of Rheumatoid Arthritis and Its Age of Onset through Modelling Genetic Risk Variants with Smoking. PLoS Genet 2013; 9: e1003808.

Scott IC, Tan R, Stahl D, Steer S, Lewis CM, Cope AP. The Protective Effect Of Alcohol On Developing Rheumatoid Arthritis: A Systematic Review And Meta-Analysis. Rheumatology (Oxford) 2013; 52: 856-67.

Other Research Papers and Review Articles

Ma MHY, **Scott IC**, Dahanayake C, Cope AP, Scott DL. Clinical And Serological Predictors Of Remission In Rheumatoid Arthritis Are Dependent On Treatment Regimes. J Rheumatol 2014; (In-Press).

Scott IC, Kingsley G, Scott DL. Can We Discontinue Synthetic Disease-Modifying Anti-Rheumatic Drugs In Rheumatoid Arthritis? Clin Exp Rheumatol 2013; 31: S4-8.

Scott IC, Lewis CM, Cope AP, Steer S. Rheumatoid Arthritis Severity: Its Underlying Prognostic Factors And How They Can Be Combined To Inform Treatment Decisions. Int J Clin Rheumatol 2013; 8: 247-263.

Scott IC, Steer S, Lewis CM, Cope AP. Precipitating And Perpetuating Factors Of Rheumatoid Arthritis Immunopathology- Linking The Triad Of Genetic Predisposition, Environmental Risk Factors And Autoimmunity To Disease Pathogenesis. Best Pract Res Clin Rheumatol 2011; 25: 447-68.

Abstracts Presented At Meetings

Scott IC, Quist J, Spain S, Tan R, Steer S, Cope AP, Lewis CM. Amino Acid Position 10 In The HLA-DRβ1 Protein Associates With Radiological Erosions In Early Active RA. Ann Rheum Dis 2014 (In-Press).

Scott IC, Walker J, Quist J, Spain S, Tan R, Okada Y, Raychaudhuri S, Steer S, Cope AP, Lewis CM. Genetic Susceptibility Variants For Rheumatoid Arthritis Do Not Associate With Radiological Progression In Early Active Disease. Rheumatology 2014 (In-Press).

Scott IC, Seegobin SD, Ma MHY, Dahanayake C, Cope AP, Scott DL, Lewis CM. Assessment of combination therapy responses in early rheumatoid arthritis defined by presence or absence of anti-citrullinated protein antibodies. Lancet 2014; 383: S95.

Scott IC, Seegobin SD, Steer S, Tan R, Forabosco P, Hinks A, Eyre S, Morgan AW, Wilson AG, Hocking LJ, Wordsworth P, Barton A, Worthington J, Cope AP, Lewis CM. Predicting the Risk of Rheumatoid Arthritis and Its Age of Onset: Modelling Genetic Risk Variants with Smoking. Rheumatology 2013; 52 (suppl 1): i43.

Scott IC, Steer S, Tan R, Forabosco P, Morgan AW, Hinks A, Thomson W, Barton A, Worthington J, Cope AP, Lewis CM. Prediction model for rheumatoid arthritis: modelling 46 genetic risk variants with smoking. Lancet 2013; S97.

Ma MHY, Dahanayake C, **Scott IC**, Kingsley GH, Cope AP, Scott DL. Serological status: a predictor of response to intensive therapy in rheumatoid arthritis. Lancet 2013; 381: S68.

Scott IC, Lewis CM, Steer S, Cope A. A Systematic Review and Meta-Analysis of Alcohol as a Protective Factor Against Rheumatoid Arthritis. Rheumatology 2012; 51: iii140-iii184.

Scott IC, Lewis CM, Steer S, Cope A. A Systematic Review and Meta-Analysis of Pregnancy as a Protective Factor Against Rheumatoid Arthritis. Ann Rheum Dis 2012; 71: A21.

CHAPTER 1. INTRODUCTION

1.1. An Overview of Rheumatoid Arthritis

1.1.1. Historical Background

Rheumatoid arthritis (RA) was first described by Augustin Landré-Beauvais, a physician in France in the 1800s. He managed several patients with severe joint pain, which in contrast to gout mainly affected the poor and women. He termed this condition "Primary Asthenic Gout" (Entezami, et al., 2011). Sir Alfred Garrod, an academic clinician in London in 1859, subsequently made the distinction of this condition from gout, based on the absence of high uric acid levels in the blood. His son, Sir Archibald Garrod, finally named the condition "Rheumatoid Arthritis" in 1890 (Entezami et al, 2011). It took several decades for the term RA to be universally recognised; it was not officially used by the American Rheumatism Association (ARA) until the 1940s.

1.1.2. Classification Criteria

As with many medical conditions, RA lacks a single pathognomonic feature. It can therefore be difficult to differentiate it from other inflammatory arthritidies; this is especially true in early disease. The diagnosis of RA relies on the presence of a combination of clinical, laboratory and radiological findings. Classification criteria have been developed, which combine these features to classify an individual as having RA for the purposes of understanding its aetiology and disease course (Fries, et al., 1994). Whilst these are often used as diagnostic criteria in clinical practice, they were not designed for this purpose as misclassification can occur. The gold standard for RA diagnosis is that of an experienced rheumatologist's opinion.

The first attempts to develop RA classification criteria were undertaken in the 1950s. The 1958 ARA criteria used a hierarchy of diagnostic certainty, which ranged from "possible" to "classical" disease (Ropes, et al., 1958). They were subsequently revised in 1987 by the American College of Rheumatology (ACR), during which the diagnostic certainty category was removed (Arnett, et al., 1988). According to the 1987 ACR criteria an individual is classified as having RA when 4 of 7 qualifying criteria are fulfilled (Table 1-1). Although the 1987 criteria have a high overall sensitivity and specificity, they were developed from patients with established

disease and therefore lack accuracy in classifying RA in its early stages (Banal, et al., 2009).

Table 1-1. 1987 ACR Classification Criteria for RA

Criteria	Definition
1. Early morning stiffness	Morning stiffness in and around the joints, lasting >1 hour before maximal improvement
2. Arthritis involving three or more joint areas	At least 3 joint areas have simultaneously had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas comprise the PIP, MCP, wrist, elbow, knee, ankle and MTP joints
3. Arthritis of the hand joints	At least 1 area swollen (as defined in criteria 2) in a wrist, MCP, or PIP joint
4. Symmetrical arthritis	Simultaneous involvement of the same joint areas on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry)
5. Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxta-articular regions, observed by a physician
6. Positive serum rheumatoid factor (RF)	Demonstration of abnormal amounts of serum RF by any method for which the result has been positive in <5% of normal control subjects
7. Radiographic evidence of RA	Radiographic changes typical of RA on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localised in or most marked adjacent to the involved joints (OA changes alone do not qualify)

To classify a patient as having RA, criteria 1-4 must have been present for at least 6 weeks and 4 or more criteria must be present. The classification of RA should not be made by these criteria alone if another systemic disease associated with arthritis is definitely present. PIP = proximal interphalangeal; MCP = metacarpophalangeal; MTP = metatarsophalangeal; OA = osteoarthritis.

Over the last few years there has been a major shift in the paradigm of RA management towards attaining an early diagnosis and initiating prompt intensive treatment (Deighton, et al., 2009). This approach improves both short-term and long-term outcomes (Boers, et al., 1997b, Choy, et al., 2008, Landewe, et al., 2002). As a result the 1987 ACR RA classification criteria have been updated by a joint working group of the ACR and European League Against Rheumatism (EULAR); these new criteria aim to identify patients with an undifferentiated arthritis that are likely to develop a chronic, erosive disease (Aletaha, et al., 2010). They classify the presence of RA based on synovitis in at least 1 joint, the absence of an alternative diagnosis

and the achievement of a total score of at least 6 points (from a possible 10) across 4 domains (Table 1-2).

Table 1-2. 2010 ACR/EULAR Classification Criteria for RA

1. Definite clinical synovitis (swelling) of ≥ 1 joint *and*
2. Synovitis not better explained by another pathology *and*
3. Score of ≥6/10 from the following 4 domains:
   i. Joint involvement
      - 1 large joint = 0 points
      - 2-10 large joints = 1 point
      - 1-3 small joints = 2 points
      - 4-10 small joints = 3 points
      - >10 joints (at least 1 small joint) = 5 points
   ii. Serology
      - Negative RF and negative ACPA = 0 points
      - Low-positive RF or low-positive ACPA = 2 points
      - High-positive RF or high-positive ACPA = 3 points
   iii. Acute-phase response
      - Normal CRP and ESR = 0 points
      - Abnormal CRP or ESR = 1 point
   iv. Duration of symptoms
      - <6 weeks = 0 points
      - ≥6 weeks = 1 point

To classify RA, patients must satisfy all three criteria. Low positive serological tests refer to values ≤3 times the upper limit of normal; high positive refers to values >3 times the upper limit of normal; RF = rheumatoid factor; ACPA = anti-citrullinated protein antibodies; ESR = erythrocyte sedimentation rate; CRP = C-reactive protein.

The 2010 criteria have an increased sensitivity compared to their 1987 predecessor. This reflects the designing working group's remit to increase the sensitivity of classification criteria in early disease. As a result, their specificity is lower, although this can be improved providing a careful exclusion of alternative diagnoses is undertaken prior to their application (Mjaavatten and Bykerk, 2013). Another key difference is their weighting towards scoring highly for the presence of RA specific autoantibodies; they therefore preferentially classify seropositive patients as having RA. In order for a seronegative patient to fulfil the 2010 criteria they need to have

more than 10 joints involved; the criteria therefore have low sensitivity for classifying seronegative RA (Kaneko, et al., 2011).

1.1.3. Epidemiology

RA is a relatively common disease. Its prevalence rate within the UK is often quoted as 1%. This is based on a historical survey of 1,236 males and 1,354 females in two Northern UK areas, which used the 1958 ARA RA Classification Criteria (Lawrence, 1961). A more recent study, undertaken within the Norfolk Arthritis Register (NOAR) and using the 1987 ACR classification criteria, suggests that the prevalence of RA in females is falling. Extrapolating NOAR data to the UK population indicates an overall RA prevalence in adults of 0.81% (1.16% in women and 0.44% in men) (Symmons, et al., 2002). The prevalence of RA increases with age and differs between genders, being approximately 3 times commoner in women. It also varies geographically; RA is substantially commoner in developed areas such as Northern Europe and Northern America, when compared with developing areas. The prevalence of RA in developing countries like South Africa, Nigeria, Indonesia, Pakistan, China, the Philippines and Argentina has been estimated at <0.5% (Kalla and Tikly, 2003). Such geographical discrepancies suggest differing environmental exposures and/or genetic susceptibility factors.

1.1.4. Clinical Features

1.1.4.1.Joint Involvement

The cardinal features of RA are synovitis and tenosynovitis. The main symptom patients report is, therefore, that of joint pain with swelling and early morning stiffness. This latter feature is attributable to circadian variations in endogenous corticosteroids.

RA has a predilection for the peripheral small joints although any joint can be affected. In most patients the proximal interphalangeal (PIP), metacarpophalangeal (MCP), thumb interphalangeal and wrist joints are involved. Joint involvement is often symmetrical. If synovitis is not suppressed with treatment, joint damage and deformity can occur. Characteristic deformities comprise ulnar deviation of the

fingers (due to MCP joint subluxation), swan-neck deformities, boutonnière deformities, and a Z-shaped thumb deformity.

1.1.4.2.Extra-Articular Manifestations

The extra-articular features of RA are protean (Table 1-3). They affect up to 40% of patients and are commoner in individuals that are seropositive, smokers or have early disability (Turesson, et al., 2003). Although the occurrence of rheumatoid vasculitis appears to be falling, extra-articular features remain a common complication of RA that are associated with an excess mortality (Myasoedova, et al., 2011).

Table 1-3. Extra-Articular Features of RA

Category	Features
Systemic	Fever
	Weight loss
	Fatigue
Dermatological	Nodules
	Small vessel vasculitis
	Palmar erythema
	Pyoderma gangrenosum
Pulmonary	Lung nodules
	Pleural effusion (exudate)
	Interstitial lung disease
Vasculitis	Cutaneous
	Systemic
Opthalmological	Episcleritis/scleritis
	Sjogren's syndrome
Cardiological	Pericarditis
	Valvular heart disease
Neurological	Cervical myelopathy
	Peripheral neuropathy
	Mononeuritis multiplex

Extra-articular features listed obtained from review by Cojocaru et al (Cojocaru, et al., 2010).

1.1.5. Current Management

Historically, RA was treated with disease-modifying anti-rheumatic drug (DMARD) monotherapy with treatment titrated slowly until disease activity was deemed to be adequately controlled or side-effects occurred. The last few years have seen a key shift in the paradigm of RA management towards early, intensive combination DMARDs and corticosteroids. These are titrated until a target of remission has been attained (Deighton et al, 2009). In refractory cases biologic agents are instituted at an early stage.

1.1.5.1.Disease-Modifying Anti-Rheumatic Drugs

DMARDs are a diverse group of drugs that are categorised together as they not only improve RA symptoms but also modify the disease course, reducing joint damage and disability.

Methotrexate is the dominant drug with National Institute for Health and Care Excellence (NICE) guidelines advocating its use in all early, active RA patients (Deighton et al, 2009). Other commonly used DMARDs comprise sulfasalazine, hydroxychloroquine and leflunomide. DMARDs that are rarely used comprise gold, penicillamine and ciclosporin. All DMARDs have potentially serious side-effects, chiefly haematological abnormalities and liver toxicity; they therefore require monitoring, usually in the form of regular blood tests.

Although methotrexate is widely considered to be the most effective DMARD, this is not clearly supported by published data. A systematic review by the Agency for Health Care Research and Quality (AHRQ) summarised the evidence for DMARD efficacy; it reported inconclusive evidence for differences in efficacy between oral methotrexate, sulfasalazine and leflunomide, although heterogeneity in methotrexate dosing across trials limited their comparability (Singh and Cameron, 2012).

1.1.5.2.Biologic Agents

In contrast to DMARDs, which have a general effect on the immune system, biologics target specific immune system components such as the pro-inflammatory

cytokines, TNF-α and IL-6 (Strand, et al., 2007). There are currently 5 different classes of biologics licensed for the treatment of RA (Table 1-4).

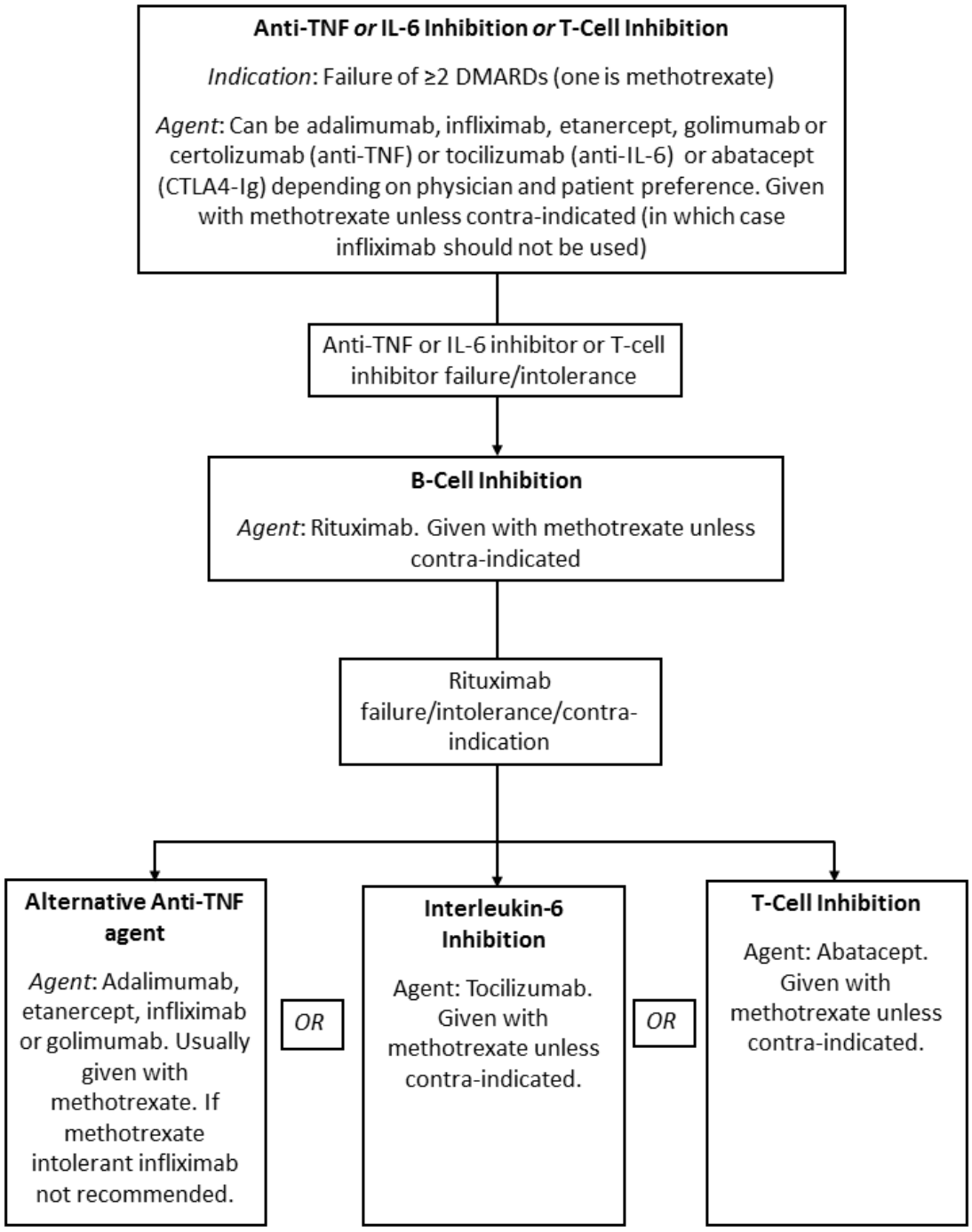
Table 1-4. Current Biologic Agents for the Treatment of RA

Biologic Class	Therapeutic Agent(s)
TNF-inhibitors	Adalimumab
	Etanercept
	Infliximab
	Certolizumab
	Golimumab
B-cell inhibition	Rituximab
T-cell inhibition	Abatacept
Interleukin-6 inhibition	Tocilizumab
*Interleukin-1 receptor antagonist**	Anakinra

** = not approved by NICE for RA management.*

Initial agents of choice for RA are TNF or IL-6 inhibitors. Their use in the UK is restricted by NICE to patients who have failed two conventional DMARDs, including methotrexate and have severely active disease defined as having a DAS28 score >5.1 on two occasions one month apart (National Institute for Health and Care Excellence, 2007). The other biologic classes are used in patients who have failed or have a contra-indication to these agents. The initial biologic in these instances is rituximab (National Institute for Health and Care Excellence, 2010); abatacept is used in cases of anti-TNF or IL-6 inhibitor and rituximab failure or contra-indication (National Institute for Health and Care Excellence, 2010). Figure 1-1 outlines the current biologic treatment paradigm for RA in the UK.

Figure 1-1. Current Biologic Treatment Pathway for RA



1.1.5.3.Small Molecule Drugs

Biologics have their limitations (Singh, et al., 2010); they are costly, require parenteral administration, have long half-lives and lack efficacy in some patients. There is therefore an ongoing need to identify novel, orally administered drugs. Pharmaceutical companies have recently focussed their

attention on identifying efficacious small-molecule agents (Stanczyk, et al., 2008). These drugs interrupt intracellular signalling by inhibiting kinases. Mitogen-Activated Protein Kinases (MAPKs), Protein Tyrosine Kinases (PTKs), Janus-Associated Kinases (JAKs) and Spleen Tyrosine Kinase have all been potential targets of interest in RA management. To date only a single kinase inhibitor, tofacitinib, has been licensed for RA management. Tofacitinib is an inhibitor of JAK; it inhibits multiple JAKs. Several large randomised controlled trials (RCTs) have evaluated tofacitinib; the results support its efficacy (Burmester, et al., 2013, Fleischmann, et al., 2012). It is approved for use in the United States of America (USA) but has not yet been approved within the UK.

1.1.6. The Concept of RA as a Syndrome

It is becoming increasingly clear that as opposed to representing a single disease, RA is in fact a clinical syndrome that spans several different disease subsets (Van Der Helm-Van Mil and Huizinga, 2008). The current main division is by autoantibody status, particularly ACPA, although other subsets probably exist. Evidence for this division stems from studies demonstrating that ACPA-positive and ACPA-negative have differing genetic (Eyre, et al., 2012) and environmental risk factors (Klareskog, et al., 2006b, Scott, et al., 2013c), phenotypes (Van Der Helm-Van Mil, et al., 2005) and treatment responses (Gottenberg, et al., 2012, Isaacs, et al., 2013, Potter, et al., 2009). It is crucial to consider this disease heterogeneity when researching and treating patients with RA.

1.1.7. Immunopathology

RA is characterised by chronic systemic and articular inflammation. Although the precise immunopathological mechanisms that underlie RA have not been completely defined this process is driven by both the innate and adaptive immune systems with T cells, B cells, macrophages, neutrophils and synovial fibroblasts playing important roles.

1.1.7.1.Macrophages

Macrophages and their precursor's monocytes act both systemically, through the production of classical RA pro-inflammatory cytokines i.e. IL-1 and TNF-α (Firestein, et al., 1990) and locally, through synovial infiltration with macrophages/monocytes being enriched in the rheumatoid synovium and destructive pannus tissue (Burmester, et al., 1997).

1.1.7.2.T Cells

T cells are well established as key components of RA immunopathology. Evidence for this stems from the strong association of RA with the shared epitope (SE) alleles encoding the major histocompatibility complex (MHC) (indicating that antigen presentation to T cells is an important process in RA), the prevalence of CD4+ T cells within rheumatoid synovium and the efficacy of the selective T cell co-stimulation modulator, abatacept (Genovese, et al., 2005). Antigen-dependent T cell responses may be important in initiating the inflammatory response during RA (Andersson, et al., 2008). They may also act independently of antigenic stimulation to perpetuate inflammation through activating monocytes/macrophages to produce pro-inflammatory cytokines (Sebbag, et al., 1997). Additionally T helper 17 cells produce IL17, which has pleiotropic effects on many RA effector cells causing inflammation and driving osteoclastogenesis and bone resorption (Kotake, et al., 1999).

1.1.7.3.B Cells

The importance of B cells in RA is highlighted by the efficacy of B cell depletion therapy with the anti-CD20 monoclonal antibody, rituximab (Cohen, et al., 2006). B cells have multiple functions in RA. They can act as antigen presenting cells (APCs) presenting antigens via MHC class II molecules to T cells activating them with consequent downstream macrophage activation and TNF-α production. They produce pro-inflammatory cytokines and chemokines directly via Toll-like receptor activation (Martinez-Gamboa, et al., 2006). B cells are responsible for autoantibody production, with RF and ACPA forming immune complexes with immune responses via Fc and complement receptors (Carroll, 2004). Additionally in many RA patients

synovial extra-follicular germinal centres develop with B lymphocytes surrounded by T cells, acting as functional ectopic germinal centres (Schroder, et al., 1996).

1.1.7.4.Synovial Fibroblasts

Fibroblast-like synoviocytes (FLS) are prevalent in RA synovium where they have a unique phenotype with aggressive and invasive properties; they drive cartilage erosion through matrix metalloproteinase production and are dominant producers of IL-6 (Bartok and Firestein, 2010).

1.2. Rheumatoid Arthritis Susceptibility Factors

1.2.1. An Overview

RA is considered to occur when genetically predisposed individuals are exposed to specific environmental risk factors. These gene-environment risks interact to trigger perturbations in the immune system, with autoantibody (RF and/or ACPA) generation in the majority of cases, followed by pro-inflammatory cytokine production and a consequent inflammatory arthritis (Scott, et al., 2011).

Over the last few decades epidemiological studies have proposed a variety of different environmental risk factors for RA. With the exception of cigarette smoking (Sugiyama, et al., 2010), their associations are often shown in case-control, but not cohort studies; their links to RA development are therefore uncertain.

Recent advances in affordable genotyping techniques have allowed the genome-wide analysis of large numbers of individuals with and without RA. The genetic architecture of RA susceptibility is therefore relatively well characterised, with 101 validated independent RA susceptibility loci identified in a recent meta-analysis of genome-wide association studies (GWASs) (Okada, et al., 2013). Recent estimates suggest that in European populations approximately half of the proportion of RA heritability has been identified, of which 15% and 36% is explained by non-MHC susceptibility SNPs and HLA alleles, respectively (Eyre et al, 2012).

1.2.2. Genetic RA Susceptibility Factors

Genetic factors dominate an individual's risk of RA. They are estimated to account for approximately two-thirds of the overall risk burden for both ACPA-positive and ACPA-negative disease (Van Der Woude, et al., 2009). The genetic architecture of ACPA-positive RA is better defined. This partly reflects the fact that existing GWAS meta-analyses have examined smaller numbers of ACPA-negative patients (Eyre et al, 2012); it also stems from the fact that ACPA-positive RA is a more easily defined phenotype, which is at less risk of misclassification.

1.2.2.1.HLA Risk

The majority of genetic risk for seropositive RA is derived from the HLA region, specifically *HLA-DRB1*. This group of alleles encode the HLA class II DRβ-chain, which plays a pivotal role in antigen presentation; it probably influences RA susceptibility by affecting the binding and presentation of arthritogenic peptides to auto-reactive CD4+ T cells (Hill, et al., 2003).

Historically, the relationship between the HLA region and seropositive RA has been explained by the presence of a group of amino acid sequences (QRRAA, RRRAA and QKRAA) spanning positions 70-74 of the HLA-DRβ1 molecule. The classical *HLA-DRB1* alleles encoding these sequences are termed the SE alleles (Gregersen, et al., 1987). Studies defining this association used immunological reagents that preferentially determined sequences on the exposed rim of the HLA-DR molecule. Using these techniques RA's association with less accessible regions of the molecule were poorly defined. Recent advances in computational methods alongside denser genome coverage on reference panels have enabled exploration of these regions through amino acid imputation. This was undertaken in a recent study by Raychaudhuri *et al*, which used HLA imputation to evaluate the association of amino acid polymorphisms across the HLA region with susceptibility to ACPA-positive RA (Raychaudhuri, et al., 2012). This novel analysis found the majority of the MHC's association with RA susceptibility was determined by polymorphisms in three amino acid positions (11, 71 and 74) in the HLA-DRβ1 protein and position 9 in HLA-B and HLA-DPβ1 proteins. All these positions are located within peptide-binding grooves; they therefore have a biologically plausible association with RA

development via antigen presentation. As only two of the five amino acid positions lie within the SE region this component of the HLA-DRβ1 molecule may not be the driving force behind seropositive RA development it was previously considered to be.

The HLA region has a more modest effect on susceptibility to seronegative RA. Although several small studies reported an association between *HLA-DRB1**03 and ACPA-negative RA (Irigoyen, et al., 2005, Verpoort, et al., 2005), this was not confirmed in larger populations (Ding, et al., 2009, Padyukov, et al., 2011). Possible explanations for this non-replication included low power (studies of ACPA-negative patients are substantially smaller than those of ACPA-positive patients (Padyukov et al, 2011)) and misclassification (chiefly ACPA-negative cases having an alternative type of inflammatory arthritis). A recent study has attempted to overcome these limitations by analysing HLA associations in large numbers of ACPA-negative patients (2,406 cases and 13,930 controls) and using statistical methodology to regress out the effects of misclassified *HLA-B27*-positive seronegative arthritis cases (Han, et al., 2014). Controlling for misclassification effects, this study demonstrated independent associations for serine and leucine at position 11 in HLA-DRβ1 (P=1.4x10^{-13}) and for aspartate at position 9 in HLA-B (P=2.7x10^{-12}) within the peptide binding grooves. These positions induced associations at *HLA-DRB1**03 (encoding serine at 11) and *HLA-B**08 (encoding aspartate at 9).

1.2.2.2.Non-HLA Risk Loci

Through collaborative efforts the number of validated non-HLA risk variants has rapidly expanded. The largest, most recent GWAS meta-analysis has identified 100 non-HLA RA susceptibility loci (Okada et al, 2013). This study combined GWASs of European and Asian ancestry patients, in order to optimise power to detect novel risk loci. Figure 1-2 shows the ORs from this meta-analysis for RA at each locus that has a replicated genome-wide significant association with susceptibility in Europeans. The largest risks are observed for *HLA-DRB1*, *PTPN22*, *ILF3*, *TYK2*, *IL20RB* and *TNFAIP3* (all of which has ORs >1.40). For most loci the effect sizes are very modest; approximately half have ORs ≤1.10.

Figure 1-2. Odds Ratios for Validated European RA Susceptibility Loci

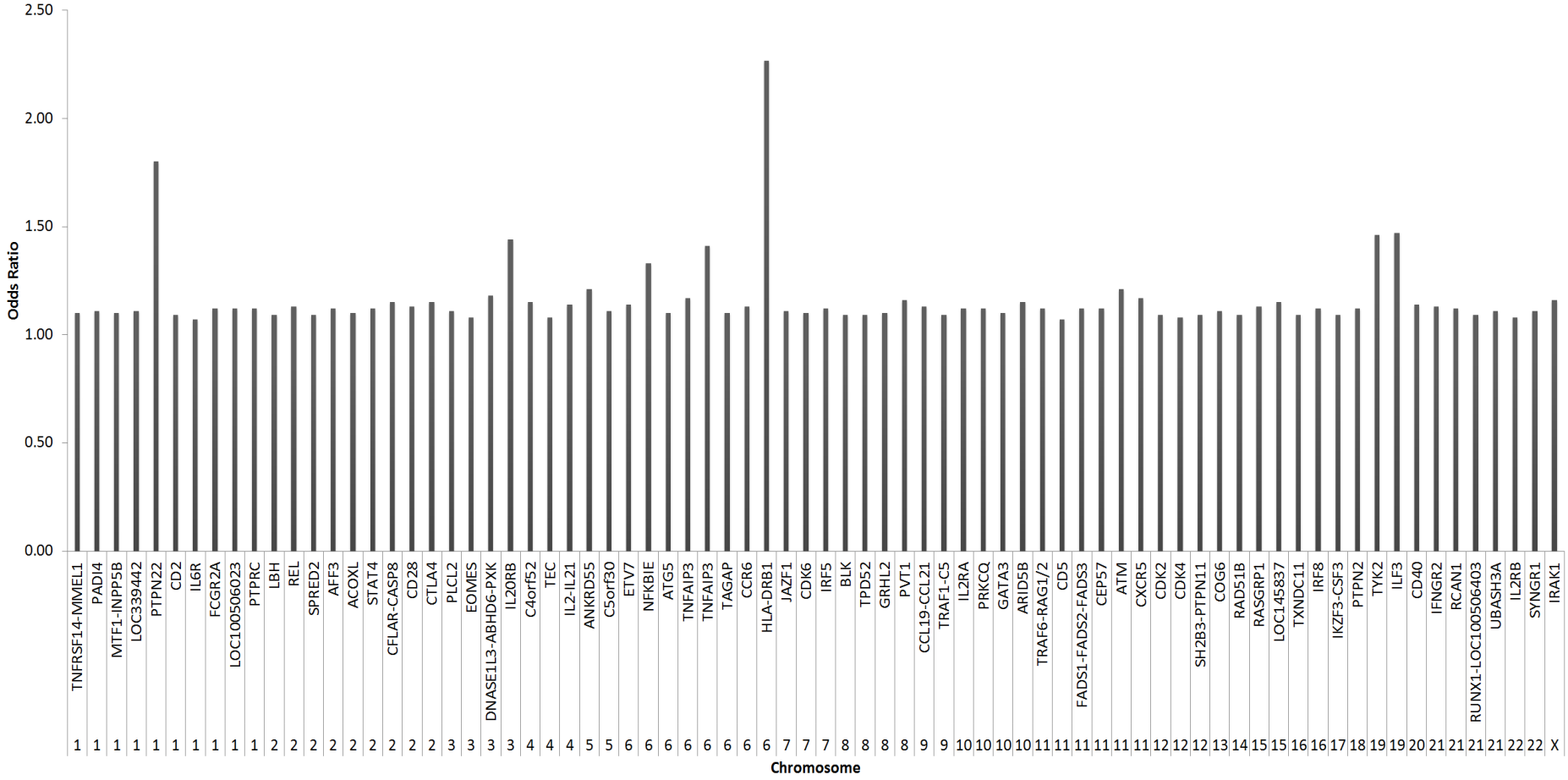


Figure adapted using data from Okada et al meta-analysis (Okada et al, 2013).

The meta-analysis by Okada *et al* did not evaluate risk stratified by serological status. The penultimate RA GWAS meta-analysis by Eyre *et al* undertook a specific analysis comparing the associations between RA subsets defined by ACPA status (Eyre et al, 2012). Of 45 non-HLA loci, approximately half had a significantly larger effect size in ACPA-positive disease, with 5 loci (*PTPN22*, *CCR6*, *CD40*, *RASGRP1* and *TAGAP*) having a markedly stronger association with this RA subset. Their findings support the concept that ACPA-positive and ACPA-negative RA have differing non-HLA genetic risk profiles.

1.2.2.3.Missing Heritability

Current estimates of the proportion of heritability explained by identified genetic risk factors suggest that approximately 50% remains unaccounted for (Eyre et al, 2012). There are several potential explanations for this. Firstly, GWASs were designed to detect associations with common variants. Rare variants could explain a substantial proportion of complex disease risk with a recent analysis using deep sequencing data suggesting that 96% of functionally important single-nucleotide variants are rare (as defined by a MAF of <0.5%) (Tennessen, et al., 2012). This appears to be the case in RA with Diogo *et al* observing an accumulation of rare nonsynonymous variants in *IL2RA* and *IL2RB* in 500 RA cases in whom deep exon sequencing of biological candidate genes in known risk loci was undertaken (Diogo, et al., 2013). Next-generation sequencing of large patient populations should identify functionally relevant rare variants; these will have larger effect sizes than those of SNPs.

Secondly, it may be that existing GWASs are underpowered to detect relevant susceptibility loci. This concept is supported by Stahl *et al*, who through the use of polygenic models demonstrated that common variants of low effect sizes accounted for a substantial proportion of RA's missing heritability (Stahl, et al., 2012).

Thirdly, current GWAS technologies do not evaluate the impact of epigenetic genome modifications on RA risk. Epigenetics is defined as "changes in gene function that are heritable and do not entail a change in the DNA sequence" (Dupont, et al., 2009). Two well characterised epigenetic mechanisms comprise post-translational histone modifications and DNA methylation; both have important

impacts on gene expression (Jaenisch and Bird, 2003). Only a few studies have evaluated the role of epigenetic mechanisms in RA susceptibility. Nakano *et al* performed a genome-wide evaluation of DNA methylation loci in fibroblast-like synoviocytes (FLS) isolated from female RA and OA patients at the time of joint replacement surgery (Nakano, et al., 2013). They demonstrated that RA and control FLS had 1,859 differentially methylated loci. Hypomethylated loci were identified in key genes relevant to RA including *STAT3* and *MAP3K5*. Liu *et al* undertook an epigenome-wide association study of 354 ACPA-positive RA cases and 337 controls. They identified 10 differentially methylated positions (DMPs), whose methylation levels could mediate genetic susceptibility in RA; 9 were within, and 1 was external to the MHC region (Liu, et al., 2013). Epigenetic studies have two potentially important confounding factors. Firstly, methylation differences may result from cellular heterogeneity within the sample material. Most DNA methylation analyses use DNA samples obtained from whole blood samples; this comprises many distinct cell populations that have been shown to vary in their methylation profiles (Reinius, et al., 2012). Secondly, as most use case-control designs they evaluate methylation profiles post-RA onset; any differential methylation patterns may therefore arise as a consequence of the disease as opposed to being causative. Liu *et al* accounted for these factors by adjusting for cell-type proportions and using mediation analyses, although the optimal strategy to exclude disease consequence methylation patterns would be to undertake epigenetic studies in a prospective cohort study, evaluating samples before and after RA onset.

Fourthly, epistasis and gene-environmental interactions may explain a significant proportion of RA heritability. This is present in seropositive RA, with the interactive effect of smoking and the SE alleles on RA risk being well established (Padyukov, et al., 2004, Pedersen, et al., 2007). Interactions between SE alleles and *PTPN22* on ACPA-positive RA risk have also been reported (Kallberg, et al., 2007).

1.2.3. Environmental Risk Factors for RA

1.2.3.1.Cigarette Smoking

Cigarette smoking is the dominant environmental risk factor for RA. It has been shown to associate with disease onset in both case-control and prospective cohort studies. Its impact appears limited to individuals with seropositive (particularly ACPA-positive) RA (Padyukov et al, 2004, Pedersen et al, 2007). Its influence is also greater in males compared with females (Sugiyama et al, 2010). A recent meta-analysis of 16 observational studies reported a summary OR for RA of 1.89 (95% CI 1.56-2.28) and 1.27 (95% CI 1.12-1.44) in male and female ever-smokers, respectively (Sugiyama et al, 2010). The risks were higher for RF-positive RA in male ever-smokers (OR 3.02; 95% CI 2.35-3.88). In female ever-smokers a non-significant association with RF-positive RA was observed (OR 1.34; 95% CI 0.99-1.80). This gender discrepancy probably reflects the fact that males smoke more heavily than females. Heavy smoking does associate with RA development in women, with a clear dose-dependent effect seen. Data from the Nurses' Health Study (NHS) indicates a linear relationship exists between RA and increasing pack years of smoking (Figure 1-3). In this cohort the age adjusted RR for RA in females with a >40 pack-year history of smoking comprised 1.99 (95% CI 1.57-2.53); in females with a 1-10 pack year history the RR comprised 1.08 (95% CI 0.86-1.36) (Costenbader, et al., 2006).

Figure 1-3. Relative Risk of RA in Women by Smoking Pack-Years

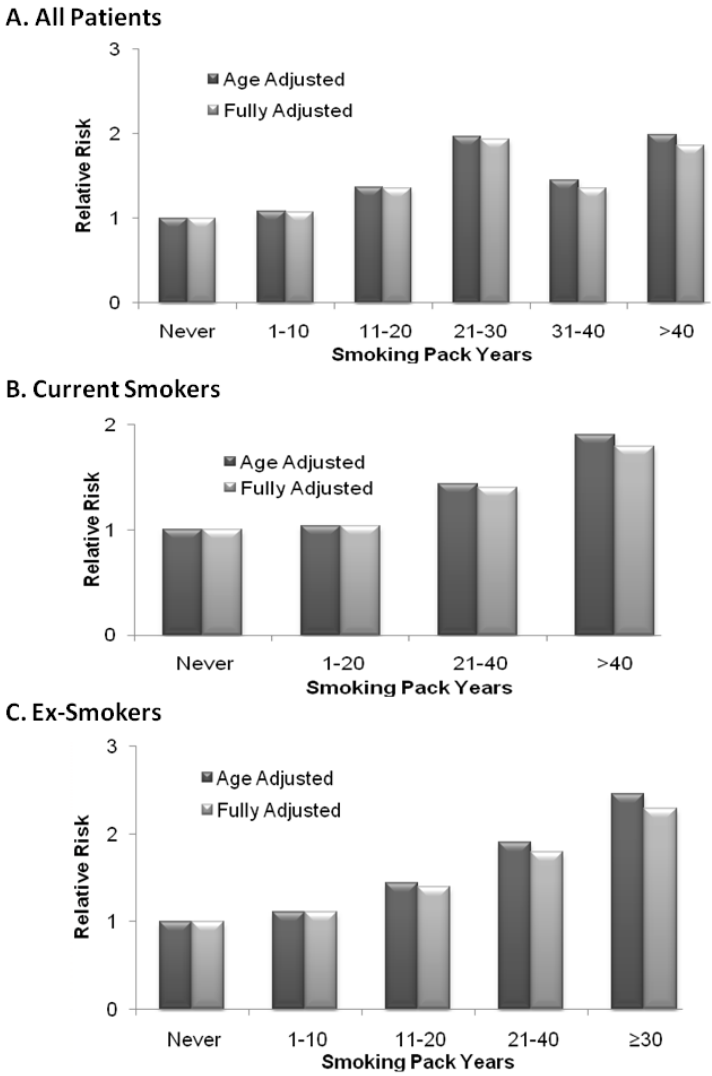


Figure adapted using data from Costenbader et al (Costenbader et al, 2006) and reproduced with permission from Scott et al (Scott et al, 2011).

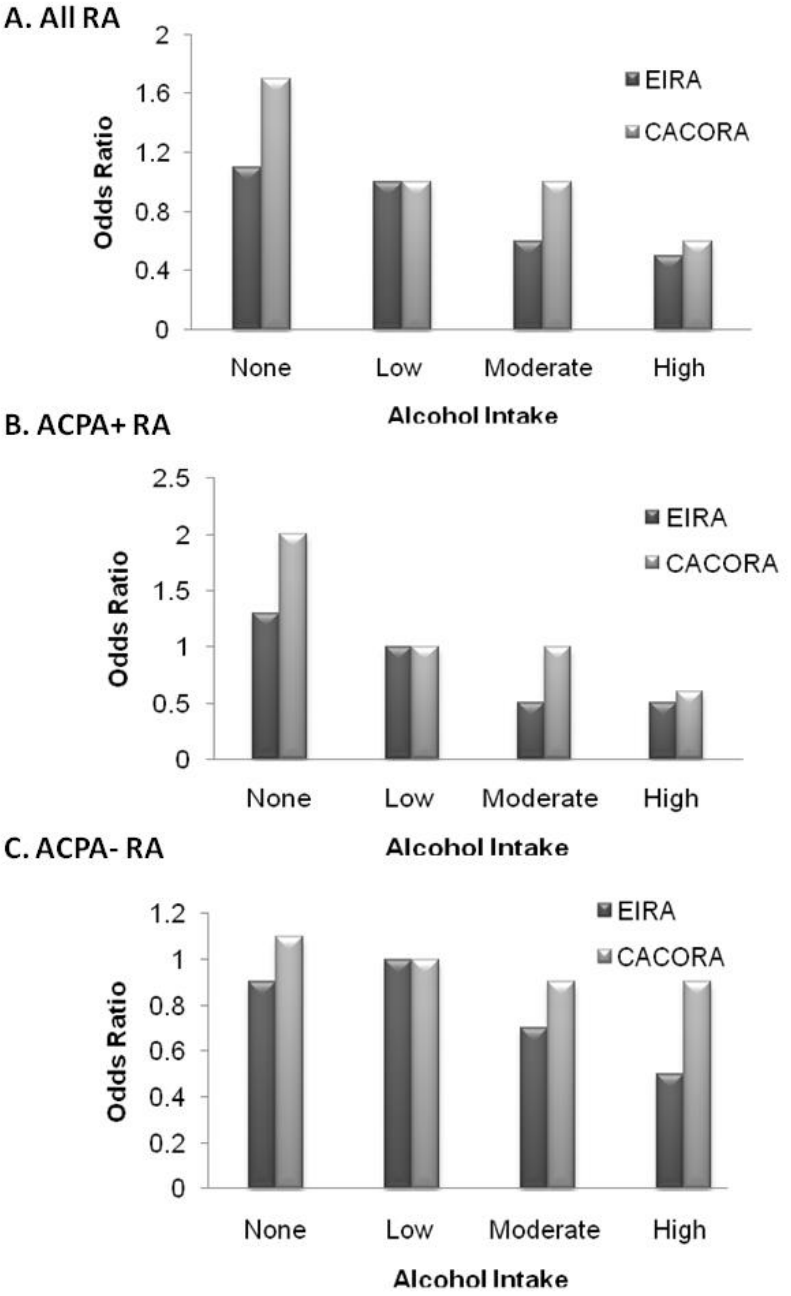
It has been proposed that smoking leads to seropositive RA through the citrullination of arginine residues; these neutrally charted neoepitopes are preferentially bound by the positively-charged pocket 4 (P4) within the SE, driving the development of ACPA and subsequently ACPA-positive RA (Wegner, et al., 2010a). This mechanism is explained in more detail later in this chapter.

1.2.3.2.Alcohol Consumption

Alcohol consumption appears to protect against RA development. Three case-control studies have shown large reductions in RA risk in alcohol drinkers compared with non-drinkers (Kallberg, et al., 2009, Maxwell, et al., 2010). Källberg *et al* evaluated data from two independent studies: the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and the Case-Control study in Rheumatoid Arthritis (CACORA) studies. In their analysis individuals consuming the highest levels of alcohol (defined as ≥5 drinks/week) had approximately half the risk of RA when compared to those consuming little/no alcohol (Kallberg et al, 2009). This protective effect was greatest for ACPA-positive disease. As with smoking, a dose-response effect was also observed (Figure 1-4). Maxwell *et al* reported the OR for RA in non-drinkers vs. individuals consuming alcohol on >10 days/month was 4.17 (95% CI 3.01-5.77) in a UK case-control study; this effect was also greater for ACPA-positive RA (Maxwell et al, 2010).

The chief criticism of these 3 case-control studies is that they asked patients to retrospectively confirm their alcohol intake after the development of RA; they were therefore subject to recall bias. Their findings have, however, recently been reproduced in 3 cohort studies, suggesting this relationship may be causal. The EPIC-2-NOAR study reported an age and sex adjusted hazard ratio (HR) of 0.83 (95% CI 0.69-0.98) for seropositive inflammatory arthritis for every 7 units of alcohol consumed/week (Lahiri, et al., 2014). Giuseppe *et al* reported an inverse association between moderate alcohol consumption and RA risk in 34,141 women in Sweden. Females consuming >3 glasses of alcohol/week in both 1987 and 1997 had a 52% reduced risk of RA compared with never-drinkers (Di Giuseppe, et al., 2012). Finally, in the NHS the pooled multivariable adjusted HR for seropositive RA in individuals consuming 5.0-9.9 grams of alcohol/day compared with non-drinkers was 0.69 (95% CI 0.50-0.95) (Lu, et al., 2014).

Figure 1-4. Effect of Alcohol on RA Risk in the EIRA and CACORA Studies



Low alcohol intake: >0 units but ≤median level of consumption by controls; Moderate alcohol intake: >median but ≤75th percentile of consumption by controls; High alcohol intake: >75th percentile of consumption by controls. All trends are significant apart from the ACPA-negative CACORA group. Figure adapted using data from Källberg et al (Kallberg et al, 2009) and reproduced with permission from Scott et al (Scott et al, 2011).

Alcohol may protect against RA through attenuation of the innate immune system. In animal models alcohol inhibited the onset of a collagen-induced inflammatory arthritis through down-regulating leukocyte migration, up-regulating testosterone secretion and reducing NF-κB activation (Jonsson, et al., 2007). Alcohol has also been shown to have anti-inflammatory effects in humans through similar mechanisms, reducing NF-κB driven inflammatory mediator production by monocytes (Mandrekar, et al., 2006), which is a key cellular pathway in RA (Dichamp, et al., 2007).

1.2.3.3.Periodontitis

Periodontitis (PD) is a destructive inflammatory disease of the teeth's supporting tissues. Its role in RA pathogenesis has received much interest for two reasons. Firstly, PD appears to be prevalent in RA patients (De Pablo, et al., 2008). Secondly, the best characterised causative organism for PD is *Porphyromonas gingivalis*; this is the only known prokaryote to possess a functional peptidylarginine deiminase enzyme, termed PPAD. PPAD has been show to citrullinate itself as well as other proteins (Wegner, et al., 2010b, Quirke, et al., 2014); it may therefore be responsible for the breakdown of immune tolerance to citrullinated autoantigens, driving ACPA formation in a similar manner to smoking (Mangat, et al., 2010).

Several studies have evaluated the link between PD and RA (Table 1-5). Most showed significant associations, although they were predominantly case-control studies, consisting of small patient numbers and using variable definitions of PD (some used dental examination outcomes like periodontal pocket depths (Abou-Raya, et al., 2008); others used self-reported histories of periodontal surgery (Arkema, et al., 2010)). The largest case-control study was a cross-sectional survey of 4,461 USA civilians. It reported that individuals with RA were more likely to be edentulous (OR 2.27; 95% CI 1.56-3.31) and have PD (OR 1.82; 95% CI 1.04-3.20) (De Pablo et al, 2008). To date the relationship between PD and RA has only been examined in a single prospective cohort study (Arkema et al, 2010). In the NHS no significant association was found between a history of periodontal surgery or tooth loss (evaluated via a self-reported questionnaire) over 12 years of follow-up. The multivariate adjusted RRs of developing RA were 1.24 (95% CI 0.83-1.83), 1.02

(95% CI 0.74-1.43) and 1.18 (95% CI 0.47-2.95) in those with previous periodontal surgery, those who had lost 1-4 teeth and those who had lost 5 or more teeth, respectively.

Table 1-5. Studies of Periodontitis as an RA Risk Factor

Study	Design	Size	Key Findings
Abou-Raya *et al* (Abou-Raya et al, 2008)	Case-control	100 cases 50 controls	72% cases and 10% controls had PD
Arkema *et al* (Arkema et al, 2010)	Prospective cohort	292 incident RA cases from 81,132 females	No association between history of periodontal surgery (RR 1.24; 95% CI 0.83-1.83) or tooth loss (RR 1.18; 95% CI 0.47-2.95) and RA
De Pablo *et al* (De Pablo et al, 2008)	Case-control	103 cases 4,358 controls	Cases more likely to have PD (OR 1.82; 95% CI 1.04-3.20)
Dissick *et al* (Dissick, et al., 2010)	Case -control	69 cases 35 controls	Moderate-severe PD commoner in cases (51%) than controls (26%) (P=0.03)
Mercado *et al* (Mercado, et al., 2001)	Case-control	65 cases 65 controls	Higher numbers of missing teeth in cases (mean 11.6) compared with controls (mean 6.7) (P<0.001)
Pischon *et al* (Pischon, et al., 2008)	Case-control	57 cases 52 controls	OR 8.05 (95% CI 2.93-22.09) for PD in cases compared with controls
Wolf *et al* (Wolff, et al., 2013)	Case-control	22 cases 22 controls	Cases had more advanced forms of PD compared with controls

PD= periodontitis; RR = relative risk; CI = confidence interval; OR = odds ratio

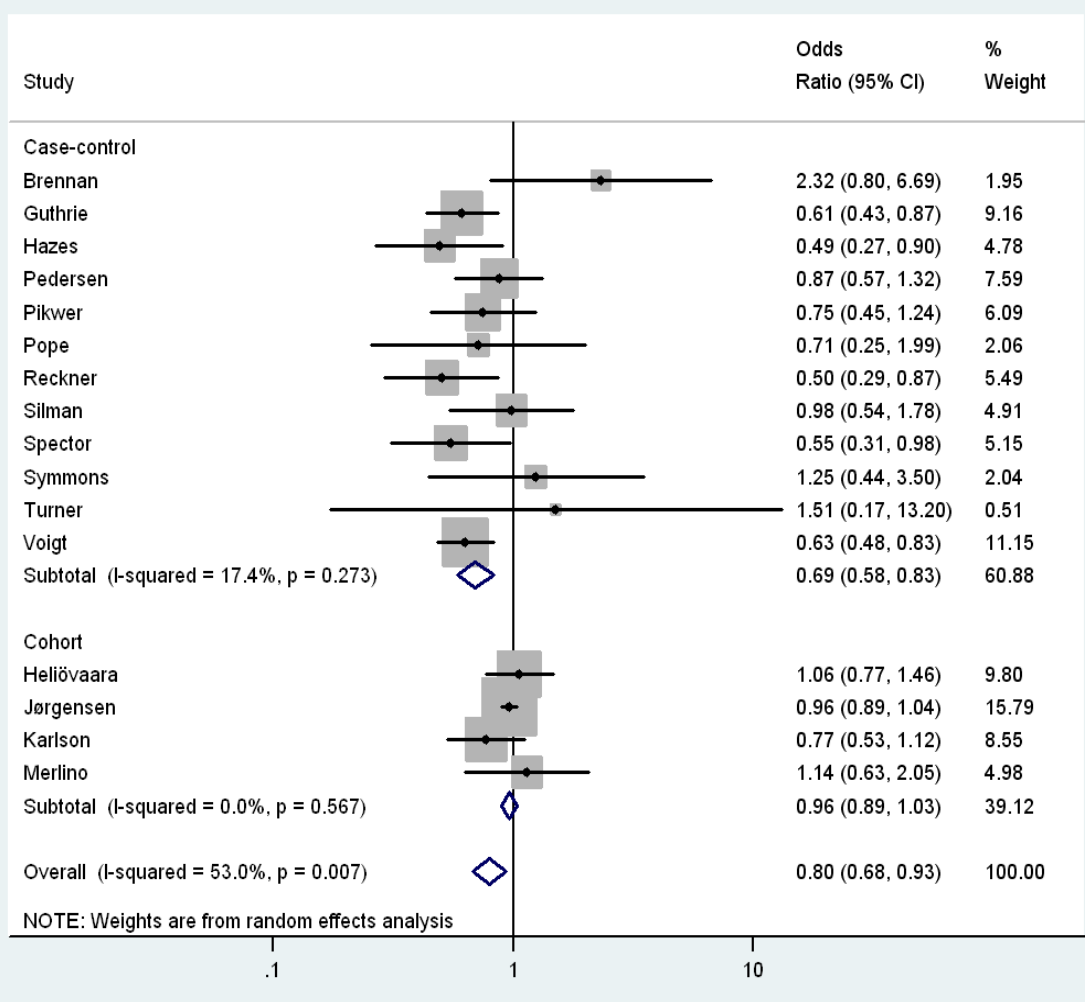
1.2.3.4.Pregnancy

RA's predilection for women has led to a marked interest in examining risk factors that females are exclusively exposed to. One such factor is pregnancy, whose relationship with RA development appears complex. In the immediate post-partum period there appears to be an increased risk of RA development; however over the longer-term a history of pregnancy appears to protect against RA.

Silman *et al* examined the relationship between the time interval from pregnancy and RA onset in 88 cases and 144 age-matched female controls (the latter group were assigned a "dummy date" for RA onset) (Silman, et al., 1992). They observed a reduced risk of RA onset during pregnancy (adjusted OR 0.30; 95% CI 0.04-2.6) and a subsequent increased risk in the first 3 months postpartum (OR 5.6; 95% CI 1.8-17.6). Similarly, Wallenius *et al* found that in 183 parous RA women the incidence rate ratio (IRR) for diagnosis during 0-24 months vs. 25-48 months post-partum was 1.73 (95% CI 1.11-2.70) (Wallenius, et al., 2010).

Multiple studies have examined the relationship between parity and RA. To date 12 case-control (Brennan and Silman, 1994, Guthrie, et al., 2010, Hazes, et al., 1990, Pedersen, et al., 2006b, Pikwer, et al., 2009, Pope, et al., 1999, Reckner Olsson, et al., 2001, Silman et al, 1992, Spector, et al., 1990, Symmons, et al., 1997, Turner and Cherry, 2000, Voigt, et al., 1994) and 4 cohort studies (Heliovaara, et al., 1995, Jorgensen, et al., 2010, Karlson, et al., 2004, Merlino, et al., 2003) have reported ORs/RRs for RA in ever- vs. never-parous women (or have provided crude data allowing for their calculation). Estimating pooled ORs for these studies using a random-effects model (due to heterogeneity) highlights a significantly reduced risk of RA in women with a history of parity (Figure 1-5); the pooled OR for RA in ever- vs. never-parous females is 0.80 (95% CI 0.68-0.93). Stratifying the meta-analysis by study design indicates this relationship is limited to case-control studies (OR 0.69; 95% CI 0.58-0.83), although a trend towards a reduced risk is also seen across cohort studies (OR 0.96; 95% CI 0.89-1.03).

Figure 1-5. Forest Plot of RA Risk in Ever- Vs. Never-Pregnant Women



Meta-analysis performed using Stata, version 10.1 (Stata Corp., College Station, TX, USA)

One proposed mechanism by which pregnancy could protect against RA is through the transfer of paternally inherited protective HLA molecules from the foetus to the mother in a process called microchimerism. Examples of such HLA molecules are those containing the "DERAA" sequence (comprising aspartic acid, glutamic acid, arginine, alanine, alanine) at *HLA-DRB1* positions 70-74; these are encoded by the *HLA-DRB1* alleles *01:03, *04:02, *11:02, *11:03, *13:01, *13:02 and *13:04 (Feitsma, et al., 2008). For this to be a convincing explanation it would require either protective HLA alleles to be transferred preferentially over susceptibility HLA alleles or males to carry a greater burden of protective HLA alleles; both of these factors are unknown but seem unlikely.

1.2.3.5.Breast-Feeding

As with pregnancy, breast-feeding is another factor that women are exclusively exposed to. Several studies have examined its impact on RA risk (Table 1-6). Of the 6 case-control studies evaluating this issue, half showed an increased risk and half a reduced risk of RA in women that had ever-breastfed. Results from the 3 cohort studies are more consistent. All showed a trend towards a reduced risk of RA with breast-feeding; this was significant in the Chinese Guangzhou Biobank Cohort (Adab, et al., 2014) and the American NHS (Karlson et al, 2004). In these two studies a dose-dependent risk reduction was observed, with significant trend tests for risk stratified by breast-feeding duration. In the Guangzhou Biobank study the adjusted OR for RA in females breast-feeding for 1-11, 12-35 and ≥36 months comprised 0.84 (95% CI 0.42-1.69), 0.58 (95% CI 0.33-1.04) and 0.54 (95% CI 0.29-1.01), respectively (*P* for trend=0.04). In the NHS the adjusted RR for RA in females breast-feeding for ≤3, 4-11, 12-23 and ≥24 months comprised 1.0 (95% CI 0.8-1.2), 0.9 (95% CI 0.7-1.1), 0.8 (95% CI 0.6-1.0) and 0.5 (95% CI 0.3-0.8), respectively (*P* for trend=0.001).

The cohort study data suggest that breast-feeding may reduce the risk of RA. Its role has some biological plausibility with higher cortisol levels observed in post-menopausal women that had previously breastfed for ≥1 year (Lankarani-Fard, et al., 2001); breast-feeding could therefore influence RA risk through perturbations in endogenous steroid levels. Another proposed mechanism is an anti-inflammatory effect from progesterone, which is mediated via an increase in the number of progesterone receptors on lymphocytes during breast-feeding (Szekeres-Bartho, et al., 2001).

Table 1-6. Studies Evaluating Breast-Feeding as an RA Risk Factor

Study	Size	OR/RR (95% CI)	Adjustment Factors
Case-Control Studies			
Berglin *et al* (Berglin, et al., 2010)	70 Cases 280 Controls	4.80 (1.43-15.87)	None
Brennan *et al* (Brennan and Silman, 1994)	60 Cases 160 Controls	2.01 (0.92-4.38)	None
Jorgensen *et al* (Jorgensen, et al., 1996)	176 Cases 145 Controls	1.62 (0.98-2.67)[a]	None
Pedersen *et al* (Pedersen et al, 2006b)	515 Cases 769 Controls	0.70 (0.35-1.39)[b]	Birth year, year RA diagnosis
Pikwer *et al* (Pikwer et al, 2009)	136 Cases 544 Controls	0.64 (0.39-1.06)[c]	Number of children
Reckner *et al* (Reckner Olsson et al, 2001)	179 Cases 259 Controls	0.40 (0.20-0.80)	Age
Cohort Studies			
Adab *et al* (Adab et al, 2014)	247 cases from 7,349 women	0.49 (0.25-0.95)[b]	Age, marital status, BMI, smoking, education, parity
Karlson *et al* (Karlson et al, 2004)	674 cases from 121,700 women	0.88 (0.77-0.99)[c]	Age, smoking, BMI, age at menarche, age at first birth, parity, OCP use, menstrual cycle regularity, HRT use
Merlino *et al* (Merlino et al, 2003)	158 cases from 31,336 women	0.94 (0.70-1.28)[b,c]	Age

a = OR calculated using crude data; b = risk assessed in parous women only; c = adjusted risks reported stratified by breast-feeding duration, which were combined into one common OR using an inverse variance fixed-effects model; BMI = body mass index, OCP = oral contraceptive pill; HRT = hormone replacement therapy.

1.2.3.6.Oral Contraceptive Pill Use

The impact of oral contraceptive pill (OCP) exposure on RA susceptibility has been extensively reviewed over the last decade. Its role was first examined in the 1970s, with a reduced incidence of RA in OCP users observed in the Royal College of General Practitioners' Oral Contraception Study (Wingrave and Kay, 1978). Since then a number of case-control and cohort studies have examined its association with

variable findings. A meta-analysis of 9 studies by Spector and Hochberg indicated that, as opposed to protecting against disease onset, OCP use may protect against the progression to a severe RA phenotype (Spector and Hochberg, 1990). Although an overall protective effect of OCP use on RA risk was observed in case-control studies, when their meta-analysis was subdivided by studies using cases enrolled from hospitals or the community different impacts on disease risk were observed. In case-control studies evaluating hospital based cases the OR for RA in OCP users was 0.49 (95% CI 0.39-0.63); in those evaluating population derived cases the OR was 0.95 (95% CI 0.78-1.16). The authors concluded that the most likely explanation for this discrepancy was that rather than preventing RA development, OCP use modified the disease process maintaining it as a mild or transient disorder.

1.2.3.7.Obesity

Adipose tissue is a highly dynamic organ, which releases a diverse mix of immune and inflammatory mediators involved in rheumatic disorders (Gomez, et al., 2011). The role of obesity in both RA and psoriatic arthritis development has gathered increasing interest.

Six studies (Table 1-7) have examined the association between obesity and RA development. Although the findings are variable their underlying methodology was highly heterogeneous: studies differed in their body mass index (BMI) definitions of obesity, the time point in an individual’s lifetime at which obesity was considered a risk factor and the method by which they captured BMI data.

Table 1-7. Studies Evaluating Obesity as an RA Risk Factor

Study	Size	Key Findings
Case-Control Studies		
Crowson *et al* (Crowson, et al., 2013)	813 Cases 813 Controls	OR 1.24 (95% CI 1.01-1.53) for BMI ≥30 vs. BMI <30
Pedersen *et al* (Pedersen et al, 2006b)	515 Cases 769 Controls	OR 1.57 (95% CI 1.01-2.44) for BMI ≥30 vs. BMI 18.5 to <25
Rodriguez *et al* (Rodriguez, et al., 2009)	559 Cases 4,234 Controls	OR 0.95 (95% CI 0.68-1.34) for BMI ≥30 vs. BMI 20 to ≤25
Symmons *et al* (Symmons et al, 1997)	165 Cases 178 Controls	OR 3.74 (95% CI 1.14-12.27) for BMI ≥30 vs. BMI <25
Cohort Studies		
Cerhan *et al* (Cerhan, et al., 2002)	158 cases from 31,336 women	RR 1.01 (95% CI 0.65-1.56) for BMI >29.2 vs. BMI <23.4
Hernandez Avila *et al* (Hernandez Avila, et al., 1990)	217 cases from 116,779 women	RR 1.1 (95% CI 0.7-1.9) for BMI >29 vs. BMI <21

BMI = body mass index; BMI in kg/m^2; OR = odds ratio; RR = relative risk.

Three case-control studies (Table 1-7) demonstrated a significantly increased risk of RA in obese individuals (Pedersen et al, 2006b, Symmons et al, 1997); one showed the risk was higher for RF/ACPA-positive RA (Crowson et al, 2013); another showed the risk was only significantly associated with ACPA-negative RA (Pedersen et al, 2006b). Cohort studies examining this issue showed no association between obesity and RA development (Cerhan et al, 2002, Hernandez Avila et al, 1990). Crowson *et al* proposed the lack of replication across studies could stem from the fact that the risk conferred by obesity is modest and the low prevalence of obesity in older studies meant they were underpowered to detect an effect (Crowson et al, 2013). However, in the absence of prospective cohort studies demonstrating a link, the association with obesity is uncertain.

1.2.3.8. Dietary Factors

A range of dietary factors have been examined for their association with RA; the results are variable. This may reflect the difficulties in capturing accurate dietary intake data, which is mainly undertaken through self-reported questionnaires.

A key example is vitamin D intake. There is a growing appreciation of the role vitamin D has in autoimmunity, with its active form producing and maintaining immunological self-tolerance (Ginanjar, et al., 2007). Analysis of data from the Iowa women's health study – a prospective cohort study of 29,368 women aged 55-69, of whom 152 developed RA over 11 years of follow-up – demonstrated an inverse relationship between vitamin D intake and RA (Merlino, et al., 2004). Individuals with the highest vitamin D intake had a RR of 0.67 (95% CI 0.44-1.00) for RA compared to individuals with the lowest intake. This relationship has not been reproduced in other datasets, like the NHS (Costenbader, et al., 2008) .

Another well evaluated dietary factor is caffeine intake. This has been assessed in 3 cohort studies. Heliövaara *et al* reported a significant association between coffee consumption and RF-positive RA in 18,981 individuals (Heliovaara, et al., 2000). Mikuls *et al* reported a significant association between decaffeinated coffee consumption and RA in 31,336 women (Mikuls, et al., 2002). Karlson *et al* reported no association between coffee consumption and RA risk amongst 83,124 women (Karlson, et al., 2003). Its link to RA development is inconsistent and therefore uncertain.

1.2.3.9.Previous Blood Transfusion

Allogenic blood transfusions have immunomodulatory effects; exposure to foreign antigens in transfused blood is associated with altered B cell populations alongside increased autoantibody production (Paglieroni, et al., 1995). UK and USA datasets have providing contrasting results for the impact of blood transfusions on RA risk. Data from NOAR indicated that blood transfusions increased the risk of RA (OR 3.58; 95% CI 1.46-8.81) (Symmons et al, 1997). The USA Iowa cohort reported that blood transfusions associated with a reduced risk of RA, although this inverse relationship was not significant (RR 0.72; 95% CI 0.48–1.08) (Cerhan et al, 2002).

The underlying reasons for these contrasting findings are uncertain; they may reflect variations in international transfusion practices, although they probably simply represent chance findings.

1.2.3.10.Socioeconomic Status

RA is commoner in lower socioeconomic populations. The EIRA study reported that individuals without a university degree had a RR of 1.5 (95% CI 1.1-1.9) for RA when compared to those with a university degree (Bengtsson, et al., 2005). Risks were greater for RF-positive RA and were mainly seen in women. The association was not explained by higher smoking rates acting as a confounder, as the RR was adjusted for age, residential area, sex and smoking. These findings were reproduced in a Danish cohort, in which education levels were inversely associated with RA risk: the multivariate OR for RA in those with the longest formal education vs. those with the lowest education level was 0.43 (95% CI 0.24-0.76) (Pedersen, et al., 2006a). Again this inverse association was predominantly seen with RF-positive RA. The probable explanation for this link is differences in environmental exposures.

1.2.4. How Gene-Environment Risk Factors Lead To RA

The underlying paradigm of RA pathogenesis is that those individuals harbouring genetic susceptibility variants are exposed to environmental risk factors. In a proportion of individuals these gene-environment risks interact to precipitate immunological changes (often characterised by autoantibody production) and arthralgia. This may progress to an unclassified arthritis followed by a fully expressed RA phenotype (Gerlag, et al., 2012).

Attempts to provide a unifying model in which the various risk factors interact to precipitate RA have been largely unsuccessful. This reflects the heterogeneous nature of RA (with different subsets probably precipitated by different factors) alongside limited characterisation of the relevant gene-environment factors involved.

Several research groups have proposed a model for the development of ACPA-positive RA. This involves interactions between two environmental risks (smoking and PD) and the *HLA-DRB1* SE alleles, which precipitate ACPA formation and

consequently ACPA-positive RA. This section describes the proposed biological model in detail.

1.2.4.1.Citrullinated Peptides

Citrulline is a non-standard amino acid that results from post-translational modification of arginine residues (Wegner et al, 2010a). This modification, termed citrullination or deimination, is facilitated by a family of peptidylarginine deiminase (PAD) enzymes in a calcium-dependent manner. The substitution of arginine for citrulline results in key changes in the structure and ionic charge of the peptide from positive to neutral, with a consequent potential for functional differences.

Although ACPAs are not completely confined to individuals with ACPA-positive RA, their presence is highly specific for this disease subset. A meta-analysis of 86 studies reported that ACPA had a sensitivity and specificity for RA of 67% and 95%, respectively (Nishimura, et al., 2007). This sensitivity can be increased by using custom arrays specific for articular peptides. Wagner *et al* demonstrated that at least 10% of RA patients testing negative for ACPAs with commercial ELISA assays, tested positive using a custom array of 16 citrullinated peptides/proteins detected in the RA synovium (Wagner, et al., 2013).

ACPA's pathological role in ACPA-positive RA is suggested by its presence in the serum of individuals many years prior to joint symptoms. Nielen *et al* analysed archived blood samples from blood donors: 41% of 79 RA patients had ACPAs pre-RA onset compared with 0.6% of controls (matched for age, sex, and donation date); the median time from the first positive ACPA test to symptom development was 4.8 years (Nielen, et al., 2004). Similarly, in another cohort of 90,000 blood donors from Sweden, which contained 83 incident cases of new onset RA, Rantapää-Dahlqvist *et al* reported that ACPAs occurred in 34% of pre-RA patients compared with 2% of matched controls (Rantapaa-Dahlqvist, et al., 2003).

1.2.4.2.Drivers of Protein Citrullination in Pre-RA Individuals

There is substantial evidence that smoking and PD – the former a definite, and the latter a probable RA risk factor – are at least partially responsible for driving protein citrullination pre-RA.

Smoking appears to increase citrullination through promoting alveolar cell PAD enzyme expression. Makrygiannakis *et al* reported substantial up-regulation of citrullinated proteins with enhanced PAD2 enzyme expression in bronchoalveolar lavage (BAL) cells from smokers compared with non-smokers (Makrygiannakis, et al., 2008). Furthermore, autoimmunity to citrulline may be promoted through increased thiocyanate ions produced by tobacco smoke metabolism (Quirke, et al., 2011). Thiocyante is metabolised to homocitrulline, which is similar in shape and structure to citrulline. Its increased presence in smokers may promote autoantibody formation with cross-reactivity to citrulline. Further evidence for the role of lung injury in driving neoepitope formation in pre-RA individuals is from a study of 105 patients with early, untreated RA and 43 healthy controls, which assessed the structural and immunological features of the lungs in relation to ACPA and smoking. High resolution computed tomography (HRCT) defined lung parenchymal changes were significantly greater in ACPA-positive compared with ACPA-negative RA cases and controls after adjusting for smoking status; additionally in ACPA-positive RA patients, ACPA levels were higher in the BAL fluid compared with the sera, suggesting the local production of ACPA in the lungs (Reynisdottir, et al., 2014).

As previously discussed, PD is prevalent in RA and its primary causative organism, *P. Gingivalis*, is the only known prokaryote to possess PPAD, which has been shown to citrullinate itself and other proteins (Wegner et al, 2010b, Quirke et al, 2014). This PPAD may drive citrullination pre-RA.

1.2.4.3.How Immune Tolerance to Citrulline May Be Breached Pre-RA

Hill and colleagues addressed the crucial issue of how immune tolerance to citrulline may be breached in pre-RA individuals (Hill et al, 2003). They examined the T cell response to citrullinated peptides in *HLA-DRB1**04:01 transgenic mice. MHC class II molecules harbouring the SE contain a specific amino acid sequence within P4 - a

positively charged area influencing binding of antigenic peptides that favours negatively charged amino acid residues. Upon citrullination the positively charged arginine is converted to citrulline, which lacks an ionic charge and has a better structural conformation with P4. Hill *et al* identified that as a result of these structural and ionic alterations citrullinated peptides bound to the MHC P4 in mice possessing the SE with a 100-fold greater affinity when compared with their arginine-containing precursors. This peptide/SE complex was subsequently presented to CD4+ T cells, activating them. The ability of citrullinated peptides to induce a T cell response has been replicated in human subjects by Feitsma *et al*, who found that naturally occurring citrullinated vimentin peptides were recognised by T cells from ACPA-positive *HLA-DRB1**04 carrying RA patients (Feitsma, et al., 2010).

The ability of ACPA to drive actual joint inflammation was subsequently shown in a mouse model (Hill, et al., 2008). *HLA-DRB1**04:01 transgenic mice were immunised with citrullinated human fibrinogen; their outcomes were compared to wild type mice. Only mice possessing the HLA transgene developed an inflammatory arthritis, which directly implicates citrullinated fibrinogen as an arthritogenic peptide in the context of SE containing MHC class II molecules.

In summary it appears probable that ACPA formation is triggered by citrullinated peptides binding with a high-affinity to SE containing MHC molecules. This peptide/SE complex is then presented to CD4+ T cells, which activate B cells driving ACPA formation.

1.2.4.4.Progression from Asymptomatic ACPA-Positivity to Clinical RA

As ACPA can predate clinical RA by many years and not all individuals with ACPA develop RA, further factors are required to trigger the shift from being an asymptomatic individual with ACPA to having established ACPA-positive RA with synovitis.

An important feature of pathogenic antibodies is that they possess fine specificity for certain antigens. It has been postulated that for ACPAs to become pathogenic and elicit articular damage they need to fully mature and increase their number of antigen

specificities. This would explain the latent period existing between ACPA formation and RA onset. Evidence for this maturation process in RA, termed "epitope-spreading", exists with Van der Woude *et al* reporting the number of citrullinated antigens recognised by ACPA increases in the time period leading up to RA onset in individuals with an undifferentiated arthritis (Van Der Woude, et al., 2010a).

Although the precise mechanisms that stimulate epitope-spreading and the onset of RA remain elusive it has been proposed that a second event such as infection or trauma occurs, which triggers a non-specific synovitis with associated citrullination within the joint (Klareskog, et al., 2006a). In healthy individuals this would resolve without sequelae, however in those possessing ACPA and T cells reactive to them this would lead to a chronic inflammatory arthritis evolving into ACPA-positive RA. ACPA mediated immune complexes could subsequently drive macrophage TNF-α production alongside other pro-inflammatory cytokine pathways (Clavel, et al., 2008).

1.2.4.5. Evidence for Relationship between Smoking, the SE and ACPA-Positive RA

If smoking were to lead to ACPA-positive RA through the preferential binding of citrullinated proteins by the SE, then one would expect to observe an interaction between smoking and the SE on the risk of ACPA-positive RA. This interaction has been confirmed in a number of well conducted case-control studies. Padyukov *et al* reported that the RR of seropositive RA was 2.4 (95% CI 1.3–4.6) in smokers with no SE alleles, 5.5 (95% CI 3.0–10.0) in current smokers with one SE allele and 15.7 (95% CI 7.2–34.2) in current smokers with two SE alleles. This dose-risk relationship is highlighted in Figure 1-6.

Figure 1-6. Smoking-Shared Epitope Interactive Effect on RA Risk

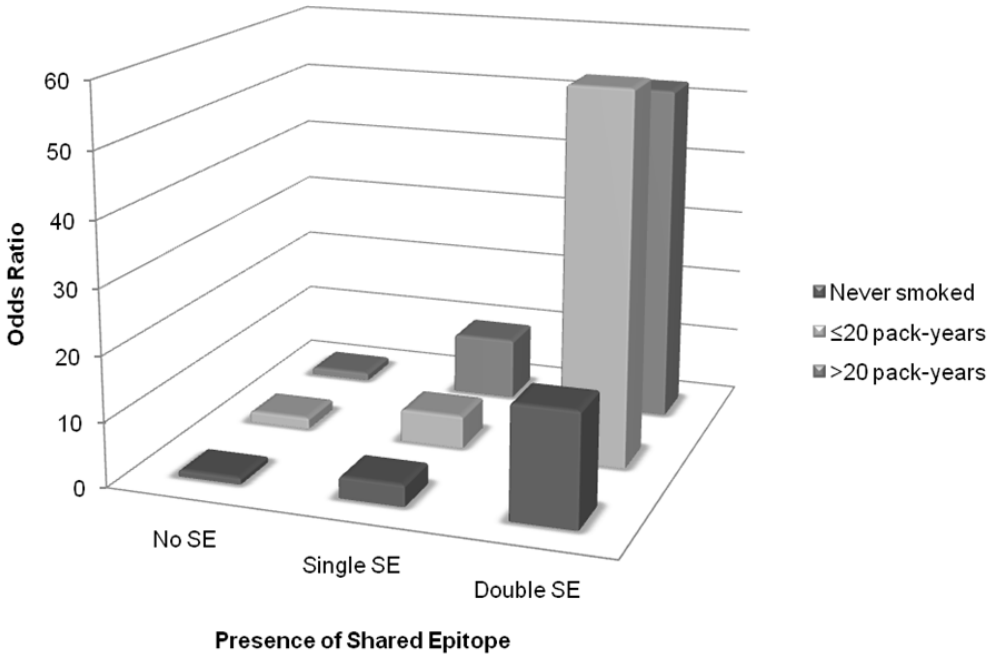


Figure adapted using data from Pedersen et al (Pedersen et al, 2007) and reproduced with permission from Scott et al (Scott et al, 2011).

Similarly, Pedersen *et al* reported that the OR for ACPA-positive RA in individuals smoking more than 20 pack-years comprised 1.22 (95% CI 0.48-3.08) in SE non-carriers, 9.66 (95% CI 4.38-21.3) in SE heterozygotes and 52.6 (95% CI 18.0-154) in SE homozygotes (Pedersen et al, 2007).

1.2.4.6.How Smoking/PD May Interact with the SE to Cause ACPA-Positive RA

In summary the following represents a biologically plausible model through smoking and PD interact with the SE alleles to drive ACPA-positive RA in some individuals (Figure 1-7).

1. Peripheral protein citrullination occurs- either through periodontal infection with *P. Gingivalis* or smoking
2. Structural and ionic changes resulting from protein citrullination enhance the binding of citrullinated self-antigens to MHC class II molecules in individuals possessing the SE
3. SE/citrullinated peptide complexes are presented by antigen presenting cells to CD4+ T cells, activating them.
4. These in turn activate B cells driving ACPA formation
5. A second factor - for example trauma or infection - precipitates synovial inflammation and the development of citrullinated proteins within the articular environment. These are targeted by ACPA.
6. ACPA-citrulline immune complexes are formed. These activate macrophages/monocytes driving inflammatory cytokine production and precipitating RA

Although this model explains the possible aetiology of some individuals with ACPA-positive RA, it fails to explain what causes this RA subset in non-smokers without PD or those individuals that don't carry SE alleles. Additionally, it fails to explain the aetiology of ACPA-negative disease.

Figure 1-7. How Gene-Environment Interactions May Cause ACPA-Positive RA

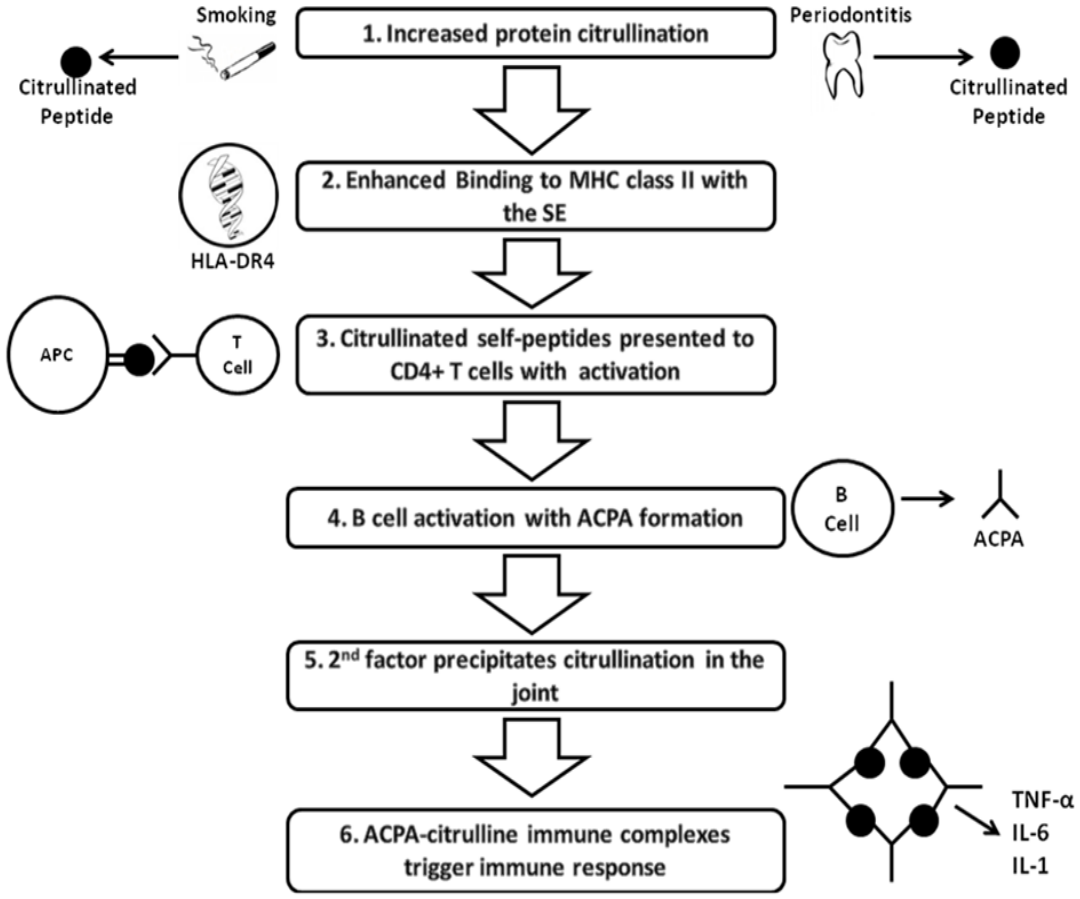


Figure reproduced with permission from Scott et al (Scott et al, 2011).

1.3. Risk Prediction Models for RA Development

1.3.1. Why These Are Needed

The high healthcare costs of RA, alongside the limitations of current RA treatments, which result in remission in only 16-42% of patients (Ma, et al., 2010) indicate that there is a key need to adopt novel management strategies. One facet of this could be instituting preventive treatments in individuals at a high-risk of developing RA. Risk prediction models are therefore needed to identify such high-risk individuals.

1.3.2. Primary versus Secondary RA Prevention

Bykerk classified RA prevention into two groups: (1) primary prevention- undertaken in individuals with genetic risks for RA in whom the pathogenic process has not yet started; (2) secondary prevention- in individuals with pre-clinical disease either at an early asymptomatic stage or in the late pre-clinical stages when symptoms are present (Bykerk, 2011).

No reliable method has been devised that can accurately identify asymptomatic individuals at a high-risk of RA from the general population. It is therefore not currently possible to perform primary prevention strategies, although as our knowledge of RA risks advance this could become an area of therapeutic potential.

Secondary prevention strategies have been instituted with some success. Patients with pre-clinical RA were identified by the presence of biological markers (like ACPA) alongside early clinical manifestations associated with progression to RA.

1.3.3. Corticosteroids for Secondary Prevention

There is some evidence that corticosteroids can attenuate RA development. Bos *et al* evaluated 83 seropositive patients with arthralgia, randomising them to receive either intramuscular (IM) dexamethasone or placebo. Corticosteroids reduced ACPA titres within one month and reduced disease activity scores at RA onset (Bos, et al., 2010a). Similarly, Verstappen and colleagues evaluated whether treating very early undifferentiated inflammatory polyarthritis (VEIA) with corticosteroids reduced future requirements for DMARDs and RA development. VEIA patients were randomised to once weekly IM 80mg depomederone injections for 3 weeks or placebo; at 12 months the number of placebo treated patients requiring DMARDs was twice that of those receiving corticosteroids (Verstappen, et al., 2010). These studies provide some support for a role of corticosteroids in RA prevention, the efficacy of which may be improved with more intensive, prolonged regimes.

1.3.4. Methotrexate for Secondary Prevention

The PRObable rheumatoid arthritis: Methotrexate versus Placebo Treatment (PROMPT) study evaluated methotrexate as a preventative treatment for RA (Van

Dongen, et al., 2007). In this RCT 110 patients with an undifferentiated arthritis (fulfilling the ARA criteria for probable RA) were treated with methotrexate or placebo. The starting dose of 15mg/week methotrexate was titrated every 3 months until the DAS28 was ≤2.4; at 12 months treatment was tapered and stopped. By the end of the study period (30 months) 40% of patients in the methotrexate group had progressed to RA compared with 53% of the placebo group (P = 0.046). Individuals treated with methotrexate also developed RA at a later time point and had less radiological progression than those treated with placebo. This trial provides some evidence for the use of methotrexate for RA prevention.

1.3.5. Biologics for Secondary Prevention

The ADJUST trial examined the impact of T cell co-stimulation modulation on the development of RA in patients with an undifferentiated arthritis or very early RA (Emery, et al., 2010). Patients were randomised to 6 months treatment with abatacept or placebo. At 2 years a non-statistically significant reduction in individuals meeting the RA classification criteria was observed in those treated with abatacept versus placebo: 46% of abatacept treated patients were classified with RA compared with 67% of the placebo group. Placebo treated patients also more frequently developed MRI structural changes.

The effect of infliximab on preventing RA development was evaluated in a small randomised study in which 17 patients with an undifferentiated arthritis of less than 12 months duration were randomised to infliximab or placebo for 14 weeks (Saleem, et al., 2008). Infliximab had no impact on progression to RA with 100% of patients in the treatment arm developing RA.

1.3.6. Published Risk Prediction Models for RA Development

In order for primary prevention to be possible risk prediction models are required that identify asymptomatic individuals at a high-risk of developing RA in the future. In order for secondary prevention to occur models are required that identify pre-RA individuals with either arthralgia or an undifferentiated arthritis that are likely to progress to a full RA phenotype. Prediction models have been developed that address both eventualities.

1.3.6.1.Methods to Assess Risk Prediction Model Accuracy

In order to place the results of published RA prediction models into context, it is important to first consider how the accuracy of these models can be assessed. A number of methods have been developed to evaluate prediction model performance. The dominant test is the area under the receiver operating characteristic curve (AUC). This is a measure of discrimination i.e. the ability of the model to discriminate a diseased from a non-diseased individual. This is established methodology in determining genetic classification test efficacy (Metz, 1978, Lu and Elston, 2008). The ROC curve plots the sensitivity (true positive rate) against 1-specificity (false positive rate) for each consecutive cut-off of the test used to define an outcome (in this case RA vs. no RA) (Steyerberg, et al., 2010). The AUC represents a summary statistic of the discriminative accuracy across the range of these cut-off values. Higher AUCs indicate better classification. An AUC >0.5 signifies some discriminative ability; a perfect classifier has an AUC of 1. It has been proposed that an AUC of 0.75-0.80 could represent an appropriate value to establish if a test may be useful for disease screening; similarly an AUC of 0.99 has been proposed as an appropriate value for a pre-symptomatic diagnostic test (Janssens, et al., 2007). These values are somewhat arbitrary and there are multiple other issues to consider before deciding whether a prediction model is suitable for use as a screening test such as those proposed by Wilson and Junger in 1968 (Andermann, et al., 2008). The AUC has a number of other limitations when applied to assessing predictive tests. Firstly, it fails to consider time, which is an important factor in prediction models whose primary goal is to estimate an individual's probability of developing a disease in the future (Cook, 2008). Secondly, AUC values are based on ranks of predicted probabilities as opposed to estimated probability values themselves; they could therefore be insensitive to differences in predicted risks between prediction models (Cook, 2008). Thirdly, for a risk factor to significantly impact on the AUC value it requires a large effect size; in the case of RA risk factors (which generally have small ORs) the AUC may not capture small discriminative improvements obtained with adding in these factors (Pepe, et al., 2004).

Several other methods are available for evaluating risk prediction model accuracy. Firstly, the overall performance of the model can be assessed by measuring the

explained variation (R^2) of the outcome captured by the model. Secondly, tests can be used to assess the calibration of the model, which is the agreement between the observed and predicted outcomes in groups of individuals e.g. if the model predicts a 50% absolute risk of RA, then the frequency of RA should be 50 in 100 individuals (Steyerberg et al, 2010). One example of a calibration test is the Hosmer-Lemeshow goodness-of-fit test (Lemeshow and Hosmer, 1982). This compares estimated observed proportions and average predicted probabilities between subgroups; a non-significant *P*-value indicates a good model fit. Thirdly, reclassification tests can be used (Pencina, et al., 2011). These are considered to be more sensitive to improvements in discrimination compared with the AUC. Reclassification tests compare the abilities of different models to correctly classify individuals into high- and low-risk categories. An example is the net reclassification improvement (NRI) test, which compares the increase in the number of individuals correctly classified as high-risk and low-risk for a disease between two models that differ by the presence of absence of a specific predictive factor (Pencina et al, 2011).

1.3.6.2.Risk Prediction Models for RA in Asymptomatic Individuals

Risk prediction models have been developed that attempt to predict the future risk of disease in asymptomatic individuals using gene-environment data for a range of immune-mediated complex disorders. For disorders without an HLA dominant genetic contribution the addition of genetic information to clinical data generally failed to improve the prediction model's performance. This is exemplified by a prediction model for type II diabetes mellitus (Talmud, et al., 2010). This report evaluated the role of adding genetic data from 20 SNPs associated with type II diabetes susceptibility to two established clinical risk models predicting disease development, the Cambridge risk score (combining data on age, sex, drug treatment, family history of type II diabetes, BMI and smoking status) and the Framingham offspring study type II diabetes risk score (combining data on age, sex, parental history of type II diabetes, BMI, high density lipoprotein cholesterol, triglycerides, fasting glucose). The addition of a genetic risk score failed to improve upon either of the clinical models discriminative or classification capabilities; the AUCs for the Cambridge risk score with and without genetic data were 0.72 (95% CI 0.69-0.76) and 0.73 (95% CI 0.69-0.76), respectively. For disorders with an HLA dominant

genetic contribution genetic data can be substantially more informative. One such disorder is coeliac disease, which has a strong association with the HLA types, DQ2.5, DQ8, and DQ2.2 (Abraham, et al., 2014). A recent risk prediction model, which devised a genetic risk score using L1-penalized support vector machine models, demonstrated high levels of discrimination across 6 European patient cohorts (AUCs 0.87 to 0.89) (Abraham et al, 2014). It is therefore likely that, owing to the major HLA contribution to disease susceptibility, such gene-environment prediction models may be effective for predicting the future risk of RA development.

Several research groups have attempted to develop models in this area with some success (Table 1-8). All have used the same modelling approach, using a weighted genetic risk score (wGRS), which is formed by multiplying the number of risk alleles for each SNP by the weight for that SNP (the weight is the natural log of the OR for each allele in published reference meta-analyses), and then taking the sum across the risk SNPs included in the model (De Jager, et al., 2009, Karlson, et al., 2010). All models have used the AUC to determine their ability to discriminate cases from controls.

Table 1-8. Existing Risk Prediction Models for RA in Asymptomatic Individuals

Study	Model Components	Cohorts Tested	AUC For Seropositive RA
Karlson *et al* (Karlson et al, 2010)	• 14 SNPs • 8 *HLA-DRB1* alleles • Age • Gender • Smoking	• NHS: 289 Ca; 481 Co • EIRA: 629 Ca; 623 Co	• NHS: 0.66 • EIRA: 0.75
Kurreeman *et al* (Kurreeman, et al., 2011)	• 29 SNPs	• 871 Ca, 1,229 Co	• 0.71
Chibnik *et al* (Chibnik, et al., 2011)	• 31 SNPs • 8 *HLA-DRB1* alleles	• 542 Ca; 551 Co	• 0.65
Yarwood *et al* (Yarwood, et al., 2013)	• 45 SNPs • 5 HLA amino acid polymorphisms • Smoking • Gender	• Discovery cohort: 11,366 Ca; 15,489 Co • Validation cohort: 2,206 Ca; 1,863 Co	• Discovery cohort: 0.80 • Validation cohort: 0.78

NHS = Nurses' Health Study; Ca = case; Co = control.

In RA this approach was first undertaken by Karlson *et al* (Karlson et al, 2010). Their model combined 22 genetic risk variants with the clinical factors smoking, age and gender. As some individuals in the general population have an increased disease risk and some a reduced risk, they divided their wGRS into 7 categories based on its Gaussian distribution in the control group and calculated the OR for seropositive RA for each group relative to the referent median (average-risk) group. Using this approach in their tested datasets (the NHS and EIRA studies) individuals in the highest risk group had an approximately 3-fold increased odds of disease relative to the average-risk group. The addition of genetic risk factors to a model using clinical factors alone increased its accuracy, improving the AUC from 0.63 to 0.75. However, despite relatively good discrimination their predicted absolute risks of RA remained low: in the highest risk group these were estimated at 0.7-1.3%. The clinical utility of their modelling was therefore limited. They subsequently extended their modelling approach, increasing the number of genetic and clinical risk factors incorporated and validating it in alternative patient populations (Karlson et al, 2010, Karlson, et al., 2013). This further increased the discriminative ability of their modelling. The most comprehensive model included 9 environmental risk factors, 34 genetic factors, 3 gene-environmental factor interactions and 3 environmental-environmental factor interactions; the AUC for this model was 0.72 in women and 0.77 in men (Karlson et al, 2013).

Yarwood *et al* used the same modelling methodology but incorporated a broader range of genetic susceptibility variants comprising 45 non-HLA loci, imputed amino acid polymorphisms in HLA-DRβ1 (positions 11, 71 and 74), HLA-DPβ1 (position 9) and HLA-B (position 9) alongside clinical factors (gender and smoking). The highest AUC for a model incorporating all these factors (tested in 1,978 cases and 1,224 controls) was 0.80 (Yarwood et al, 2013). Their use of HLA amino acid polymorphisms improved prediction when compared to a model incorporating *HLA-DRB1* alleles.

1.3.6.3.Risk Prediction Models for RA in Seropositive Arthralgic Patients

Patients with arthralgia and RF or ACPA are at a substantially increased risk of future RA. Not all of these individuals will, however, progress to develop an

inflammatory arthritis (Bos, et al., 2010b). Van de Stadt *et al* recently developed a prediction model for progression to arthritis in seropositive arthralgia patients (Van De Stadt, et al., 2013). This model was developed in 300 seropositive (IgM-RF or ACPA positive) patients with arthralgia but without synovitis. Variables associated with remission were identified by Cox proportional hazard analysis. The prediction model consisted of 9 variables, comprising a positive family history of RA (in a first degree family member), alcohol abstinence, symptom duration <12 months, presence of intermittent symptoms, arthralgia in the upper and lower extremities, visual analogue scale pain ≥50, presence of EMS lasting ≥1 hour, a self-reported history of swollen joints and antibody status (4 categories ranging from seronegative to positive for both IgM-RF and ACPA). This model was subsequently validated in 74 seropositive arthralgic patients recruited from the same rheumatology unit. The AUC value was 0.82 for the development of arthritis at 5 years. Patients could be categorised into three risk categories, comprising low-risk, intermediate-risk and high-risk. Using the low-risk group as a reference, the intermediate-risk group had a HR of 4.52 (95% CI 2.42–8.77) and the high-risk group had a HR of 14.86 (95% CI 8.40–28.32) for developing an arthritis. Whilst this model appears promising the fraction of explained variation captured by Nagelkerke's R^2 was only 0.31, indicating that 69% of the variance was explained by unidentified factors. Additionally, the model was validated internally in patients recruited from the same unit; it therefore requires evaluation in other external patient populations to determine its generalisability.

1.3.6.4.RA Risk Prediction Models in Early Undifferentiated Arthritis Patients

The Leiden Prediction Rule is an established model for predicting which individuals with an early undifferentiated arthritis are likely to progress to a full RA phenotype (Van Der Helm-Van Mil, et al., 2007). This prediction model was developed in 570 patients presenting with an undifferentiated arthritis that were followed-up for 12 months. Clinical characteristics associated with RA development were identified through logistic regression. Their prediction rule comprised 9 clinical variables: sex, age, symptom localisation, EMS, TJC, SJC, CRP level, RF and ACPA-positivity. This provided a score ranging from 0-14. When cut-off values of 5 and 9 were chosen, 97% of undifferentiated arthritis patients with a score ≤5 did not progress to

RA and 84% of patients with a score ≥9.0 progressed to RA. Although the AUC in the derivation cohort was high (0.89), Nagelkerke's R^2 was 0.57, indicating that 33% of variance was explained by other factors. This model has been validated in several other patient populations (UK, German and Dutch cohorts) with similar discriminative abilities demonstrated (AUCs 0.82-0.95) (Van Der Helm-Van Mil, et al., 2008). Whilst this model performed well at identifying undifferentiated arthritis patients at a high and low-risk of progressing to RA, it was unable to adequately quantify the risk of RA in those individuals with moderate scores of between 6 and 8 (Van Der Helm-Van Mil et al, 2008). This represents approximately one quarter of the patients in the validation cohorts. Further work is required to identify alternative predictors, which explain the missing variance, in order for this model to be used more readily in clinical care.

1.4. Predicting Rheumatoid Arthritis Severity

1.4.1. Defining Severe Disease

In order to identify predictors of RA severity, the first consideration is what represents severe disease. In clinical practice this is often considered to be that in which DAS28 scores are persistently above 5.1 (Prevoo, et al., 1995). Another important component of severe RA is the development of disability, which is often recorded using the Health Assessment Questionnaire Disability Index (HAQ-DI). This self-reported assessment evaluates functional ability using 20 questions spanning 8 categories; scores of 0-1, 1-2 and 2-3 are considered to represent mild-moderate, moderate-severe and severe-very severe disability, respectively (Bruce and Fries, 2003). Most studies evaluating genetic predictors of severe RA have used the extent or progression of radiological damage to define poor prognosis disease. The two main radiological outcomes in RA comprise the Sharp/van der Heijde Score (SvHS) and the modified Larsen score (Boini and Guillemin, 2001). The SvHS evaluates erosions and joint-space narrowing in 44 and 42 joints respectively alongside joint subluxation; the total score ranges 0-448. The modified Larsen score evaluates changes of erosion and joint destruction in the hands, wrists and feet providing a total score ranging from 0-200. The main benefit of using radiological scores over other outcome measures is that they generally deteriorate over time. Although correlated with other outcomes like DAS28 and HAQ scores, these latter

measures have more variable courses over time (Drossaers-Bakker, et al., 1999). It is therefore easier to detect longitudinal effects of genotypes on SvHS and Larsen scores.

1.4.2. Serological Predictors of RA Severity

The role of RF as a predictor of disease severity is well established, with individuals that are seropositive for RF consistently having higher rates of joint damage and extra-articular manifestations. This is particularly true of the IgA-RF isotype, which is often reported to have a stronger association with severe disease when compared to IgM- and IgG-RF (Jonsson and Valdimarsson, 1998). In one longitudinal observational study of 135 women with early RA, whilst all three RF isotypes significantly associated with radiological progression and higher SJCs, IgA-RF titres had the strongest correlation (Van Zeben, et al., 1992). Other studies have also shown greater associations between IgA-RF and radiological erosions (Teitsson, et al., 1984, Brik, et al., 1990) and extra-articular manifestations (Jonsson, et al., 1995) when compared to other RF isotypes. The prognostic value of ACPA is also well described. In one cohort study of 93 early RA patients identified among Swedish blood donors the presence of ACPA prior to and at disease onset significantly associated with radiological outcomes (Berglin, et al., 2006). The baseline and two year Larsen scores in cases positive for ACPA pre-disease onset were 8 and 14, respectively; for individuals negative for ACPA pre-disease onset they were 5 and 9, respectively. These differences were statistically significant ($p<0.001$) at both time points. ACPA also predicts longer-term radiological damage. Lindqvist *et al* demonstrated this in 183 RA cases followed up for 10 or more years (Lindqvist, et al., 2005). In linear regression analyses Larsen scores at 10 years significantly associated with ACPA and CRP levels, which accounted for 32% of the variance in the score.

1.4.3. Environmental and Epidemiological Risk Factors for RA Severity

A variety of environmental and epidemiological factors have been linked with RA severity (Table 1-9). Interestingly, there is substantial overlap between RA susceptibility and severity factors, with smoking and alcohol abstinence both associating with RA development and severity.

Table 1-9. Environmental and Epidemiological Prognostic Factors in RA

Risk Factor	Study	Size	Type	Severity Outcome(s)	Main Findings
Smoking	Másdóttir *et al* (Masdottir, et al., 2000)	63 Ca	Cross-sectional	Nodules, X-ray score, joint counts, HAQ	Smoking associated with nodules, higher x-ray scores, higher HAQ scores
	Manfredsdottir *et al* (Manfredsdottir, et al., 2006)	100 Ca	Longitudinal	Joint counts, pain, CRP, X-ray score	Current smokers had highest joint counts.
Alcohol	Maxwell *et al* (Maxwell et al, 2010)	873 Ca	Cross-sectional	X-ray score, DAS28-CRP, HAQ, pain	Lower x-ray scores, DAS28-CRP, CRP, HAQ and pain as alcohol intake increased.
	Nissen *et al* (Nissen, et al., 2010)	2,908 Ca	Longitudinal	X-ray score, HAQ	Trend for less x-ray progression in alcohol drinkers: progression 0.99% (95% CI 0.89–1.09) in drinkers; 1.13% (95% CI 1.01–1.26) in non-drinkers.
Periodontitis	Abou-Raya *et al* (Abou-Raya et al, 2008)	100 Ca	Cross-sectional	DAS28, HAQ, X-ray score	Periodontitis severity correlated with DAS28 score, ESR and CRP.
	Mercado *et al* (Mercado et al, 2001)	65 Ca	Cross-sectional	Joint counts, physician global, ESR/CRP, HAQ	Periodontitis severity associated with higher joint counts, HAQ and CRP/ESR levels.
Gender	Jawaheer *et al* (Jawaheer, et al., 2010)	292 Ca	Longitudinal	DAS28, HAQ, pain, physician global, CRP, X-ray score	Females had worse DAS28, global and joint count progression
	Ahlmén *et al* (Ahlmen, et al., 2010)	549 Ca	Longitudinal	DAS28, HAQ, X-ray score	Females had higher DAS28 and HAQ scores.
Social Deprivation	McEntegart *et al* (Mcentegart, et al., 1997)	814 Ca	Longitudinal	Pain, articular index, ESR, CRP, HAQ	Deprivation associated with higher HAQ
	ERAS Study Group (Eras Study Group, 2000)	869 Ca	Longitudinal	Joint counts, HAQ, Pain, ESR, X-ray score	Deprivation associated with higher HAQ and joint counts.

SJC = swollen joint count, HAQ = Health Assessment Questionnaire, TJC = tender joint count, VAS = visual analogue scale, Ca = case, Co = control, OR = odds ratio, SvHS = Sharp/van der Heijde Score, table adapted with permission from Scott et al (Scott, et al., 2013a).

1.4.3.1.Smoking

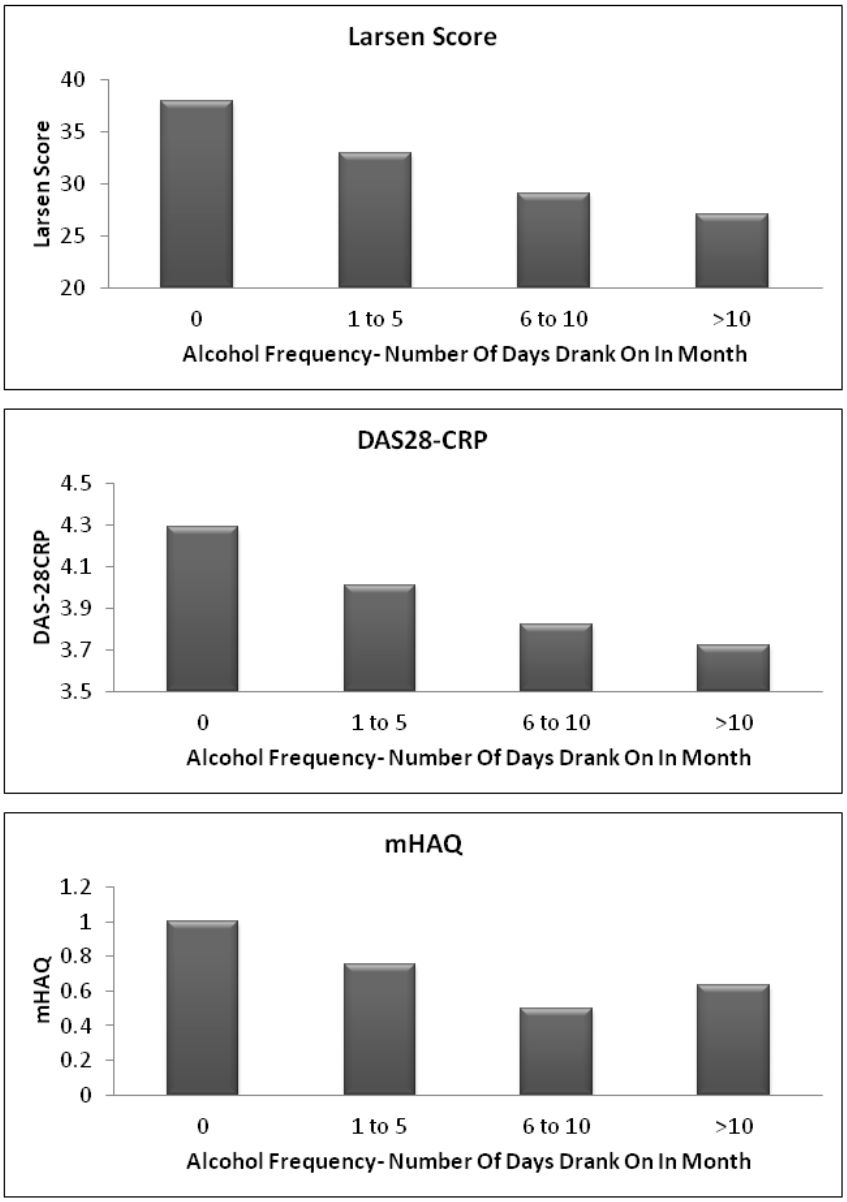
There is some evidence that smoking influences the natural history of RA. In one prospective study of 100 early RA patients followed up for 24 months the SJC, TJC and pain visual analogue scale (VAS) scores were all significantly higher in smokers when compared to non-smokers (Manfredsdottir et al, 2006). The SJC at 6 months also significantly associated with smoking status, with current smoking increasing the number of swollen joints by at least 3 on average in a regression model after the elimination of non-significant variables. Another observational study of 63 women with advanced RA (average disease duration 13.7 years) showed that heavy smoking (defined as ≧20 pack-years) significantly associated with the presence of nodules, higher radiological damage rates and higher HAQ scores when compared with smokers of <20 pack-years or never-smokers (Masdottir et al, 2000). The impact of smoking is confounded by its effect on ACPA formation, which may mediate any effect it has on disease severity (Klareskog et al, 2006b).

1.4.3.2.Alcohol Consumption

There is some evidence that increasing alcohol intake associates with a less severe disease course. In one study of 873 erosive RA cases, more frequent alcohol consumption correlated significantly with lower DAS28-CRP, Larsen and mHAQ scores (Maxwell et al, 2010). These trends are shown in Figure 1-8. The median DAS28-CRP, Larsen and mHAQ scores in individuals drinking no alcohol in the month prior to assessment comprised 4.29, 38 and 1.0 respectively; these scores in individuals drinking on more than 10 days in the month prior to assessment comprised 3.72, 27 and 0.63. All of these differences were statistically significant (P<0.05) when evaluated by trend tests across alcohol intake categories. A protective effect of alcohol intake on radiographic progression was also demonstrated in a large Swiss observational study evaluating 2,908 RA cases nested within a national database of RA patients (Nissen et al, 2010). This study evaluated the impact of drinking alcohol on the progression of X-ray damage, scored according to the Ratingen method (Rau, et al., 1998). It found that in a model adjusting for multiple variables (comprising baseline radiological damage scores, DAS28, HAQ, presence of RF, sex, age, disease duration, tobacco smoking, education level and medications) radiographic damage at 12 months had progressed by an average of 0.99% (95% CI

0.89-1.09) in drinkers and 1.13% (95% CI 1.01-1.26) in non-drinkers. Interestingly, as with the beneficial effects of drinking on cardiovascular disease, a J-shaped dose-response effect was seen with occasional and daily alcohol consumers having less radiographic progression at 12 months compared with non-drinkers and heavy drinkers. Taken together these studies provide some evidence that alcohol consumption may attenuate the inflammatory process in RA.

Figure 1-8. Relationship between Alcohol Consumption and RA Outcomes



P-values for trend tests are all <0.0001; figure adapted using data from Maxwell et al (Maxwell et al, 2010) and reproduced with permission from Scott et al (Scott et al, 2013a).

1.4.3.3.Periodontitis

The relationship between PD and RA outcomes was evaluated in a cross-sectional study of 100 patients with active RA. Significant correlations between PD severity and DAS28 scores (P<0.001), ESR (P<0.005) and high sensitivity CRP levels (p<0.003) were reported (Abou-Raya et al, 2008). Another small observational study of 65 RA patients found that individuals with moderate-severe PD had significantly more swollen joints, higher HAQ scores and higher CRP levels when compared to patients with no or mild PD (Mercado et al, 2001). As with smoking any potential effect of PD on RA outcomes could be mediated by ACPA. Further work is required with large longitudinal studies to establish this relationship, and to explore the impact of PD treatment on disease severity.

1.4.3.4.Social Deprivation

Several studies have highlighted that individuals from socially deprived areas have poorer disease outcomes (Vliet Vlieland, et al., 1994, Mcentegart et al, 1997). This association was evaluated in 869 patients from the Early Rheumatoid Arthritis Study (ERAS), which is a large prospective cohort study of individuals with RA of less than 2 years duration (Eras Study Group, 2000). The authors reported that the Carstairs score (a composite score of male unemployment, social class, overcrowding and car access that represents an index of deprivation) was associated with more severe disease at presentation, as reflected by HAQ and joint scores; this association persisted and remained after 3 years of follow-up. The precise underlying mechanism for this association is unclear; it may represent an association between low socioeconomic status and lifestyle factors like smoking.

1.4.3.5.Gender

Gender differences in RA are well described, with the incidence of RA greater amongst women compared with men (Myasoedova, et al., 2010). There is also evidence that RA outcomes are worse in females. Jawaheer *et al* found that in a longitudinal prospective study of 225 women and 67 men with seropositive early DMARD-naive RA, women had worse disease progression over 2 years as reflected by DAS28 scores, physician global scores and TJCs; this was in spite of similar

treatments (Jawaheer et al, 2010). Men were also more likely to attain remission. Similarly, the Swedish BARFOT study reported that women had significantly higher DAS28 and HAQ scores compared with males at all time-points over a 5 year follow-up period; the authors attributed this DAS28 discrepancy to a higher number of tender joints and general health scores in women compared with men (Ahlmen et al, 2010). Other studies have reported similar female gender influences on RA progression (Kuiper, et al., 2001).

1.4.4. Evidence for a Genetic Component to Radiological Damage in RA

In contrast to the identification of genetic susceptibility variants for RA there is substantially less information on which genetic markers influence RA severity. The dominant reason for this is a lack of adequately sized cohorts containing detailed genotypic and longitudinal disease outcome data. Studies evaluating this topic have generally been of limited power to detect genome-wide significant SNPs and have instead mainly relied on candidate gene approaches for the identification of relevant loci. Despite these problems there is accumulating evidence that genetics play an important role in determining radiological progression in RA. A twin study found that the variance in radiographic joint destruction was highest in unrelated patients, followed by dizygotic and finally monozygotic twins (Van Der Helm-Van Mil, et al., 2006). A more recent study has replicated the association between relatedness and radiological damage in 325 Icelandic patients with RA; this study quantified the heritability of radiological joint destruction to be between 45 to 58% (Knevel, et al., 2012b).

1.4.5. Genetics of Radiological Damage: Candidate Gene Studies

1.4.5.1.*HLA-DRB1* Alleles

HLA-DRB1 alleles, in particular those encoding the SE, have been linked to a more severe RA phenotype (Gonzalez-Gay, et al., 2002). Wagner *et al* found that in a prospective study of 55 individuals with early RA, SE carriers had an OR for erosive disease of 13.75 (P=0.00083). A meta-analysis of 3,240 RA patients demonstrated a significant association between the SE (2 or 1 versus 0 SE alleles) and erosions (OR 2.0; 95% CI 1.8–2.2) (Gorman, et al., 2004).

More recent studies evaluating this relationship have employed the classification system for *HLA-DRB1* alleles proposed by du Montcel *et al* (Du Montcel, et al., 2005). This broadly divides *HLA-DRB1* alleles into 2 groups: S alleles and X alleles, which have or do not have the RAA sequence at position 72-74, respectively. Some S alleles i.e. S_2 (containing *04:01) are associated with an increased disease risk; other S alleles i.e. S_1 alleles (containing *13:01) are associated with a reduced disease risk (Barnetche, et al., 2008). Using this classification system one observational study of 962 RA cases found that the carriage of S_2 alleles significantly correlated with higher Larsen scores with the median (IQR) Larsen score for individuals carrying one S_2 copy comprising 29 (8-61) and for those carrying two S_2 copies comprising 41 (16-73) (Mewar, et al., 2008). Carriage of S_1 alleles was associated with less radiological damage (P=0.011). Similar findings come from a prospective longitudinal study of 144 French Caucasian early RA patients in which individuals carrying S_2 alleles had greater radiographic damage progression compared with non-carriers (P=0.004) and individuals carrying S_{3D} alleles had significantly less radiographic damage progression compared with non-carriers (P<0.0001) (Gourraud, et al., 2006). In both instances significant gene-dose effects were observed.

It therefore appears that, as with disease susceptibility, some *HLA-DRB1* alleles are risk factors for, and some protect against, radiological damage in RA. As with smoking this association could be driven by ACPA.

1.4.5.2.PTPN22

The evidence that *PTPN22*, the dominant non-MHC RA susceptibility allele, contributes to radiological damage in RA is limited. In a cross-sectional study of 964 RA cases, Marinou *et al* reported a trend towards higher rates of X-ray damage in RA patients carrying the *PTPN22* minor allele compared to those without it. Median modified Larsen scores for individuals that had zero, one or two minor allele copies comprised 25.5, 33.0 and 50.0, respectively (Marinou, et al., 2007). This finding was, however, only of borderline statistical significance (P=0.04) and other studies have failed to replicate it such as the Brigham RA Sequential Study (BRASS) in which the

adjusted OR (95% CI) for an erosive phenotype in *PTPN22* T allele carriers was 1.14 (0.77-1.71) (Karlson, et al., 2008).

1.4.5.3.C5orf30

Teare *et al* recently demonstrated an allele-dose association between the RA risk variant SNP, rs26232 in the *C5orf30* loci and the extent of radiological damage in RA (Teare, et al., 2013). In this study an allele-dose association between modified Larsen/SvHS scores and the number of minor alleles was demonstrated in two UK cohorts (the Genetics of RA (GORA) study and Yorkshire Early Arthritis Register (YEAR)) using a zero-inflated negative binomial model and a Dutch cohort (the Leiden Early Arthritis Clinic (EAC) cohort), using a linear mixed-effects model. A fixed-effects meta-analysis of the incidence rate ratios (IRRs) across the 3 cohorts revealed a pooled IRR of 0.90 (95% CI 0.84-0.96) per minor allele. This can be considered as representing a reduction in radiological scores of 11% per minor allele copy. Although there is some evidence linking *C5orf30* to immune function, it is unknown how it could exert an effect on radiological damage at present.

1.4.5.4.IL1B and IL1RN

Cantagrel *et al* evaluated the relationship between two polymorphisms in the *IL1B* (interleukin 1, beta) gene and one polymorphism in the *IL1RN* (interleukin 1 receptor antagonist) gene amongst 108 patients with early RA followed up for 2 years (Cantagrel, et al., 1999). Although none independently associated with the development of erosions at 2 years, when *IL1B* exon 5 allele E2 carriage was combined with the presence of SE alleles an increased risk of erosive disease was observed: the OR (95% CI) for erosions was 8.20 (2.59-25.84). Buchs *et al* also examined the association between radiological damage and polymorphisms in the *IL1B* gene (within the promoter region at -511 and in exon V at +3954) and the *IL1RN* gene (in exon 2 at position +2018) amongst 297 RA cases (Buchs, et al., 2001). They found a significant relationship between destructive RA and the carriage of the rare *IL1B* (+3954) allele 2; carriers of this polymorphism also had higher ESR levels compared to non-carriers. The association of the exon 5 +3953 A2 allele with higher DAS28 scores and ESR levels was demonstrated in a smaller study of 93 RA patients (Pawlik, et al., 2005).

1.4.5.5.IL6

An association between an *IL6* tagging SNP and radiographic severity was reported by Marinou *et al* (Marinou et al, 2007). In this cross-sectional evaluation of 964 RA cases the SNP, rs1800795, which tags the promoter region of the *IL6* gene and is often referred to as the "-174" polymorphism, was significantly associated with radiological damage in seropositive RA. The modified Larsen scores in ACPA-positive RA risk allele non-carriers, heterozygotes and homozygotes comprised 29, 32 and 41 (trend test *P*-value=0.004). This finding has not, however, been validated in other cohorts and the prognostic relevance of *IL6* polymorphisms are uncertain.

1.4.5.6.IL10

Marinou *et al* also reported a significant association between a polymorphism in *IL10* -592C (using the tagging SNP, rs1800872) that is specific for erosive damage in ACPA-negative RA (Marinou et al, 2007). In this study ACPA-negative individuals that were homozygous for the risk allele had more severe radiographic damage compared with non-carriers/heterozygous individuals (pooled due to small numbers). The median modified Larsen score comprised 6.0 for non-carriers/heterozygotes and 16.0 for homozygotes (*P*-value for trend=0.002). Another polymorphism in the *IL10* locus was shown to influence the rate of radiological progression in a cohort of 91 patients in the Netherlands (Huizinga, et al., 2000). Although this study did not subdivide their analysis by ACPA-status the presence of the *IL10* -1082GG genotype was associated with significantly greater increases in Sharp scores at 3 and 6 years when compared to individuals with the -1082AA genotype.

1.4.5.7.TRAF1/C5

TRAF1 encodes an intracellular protein member of the tumour necrosis factor (TNF) receptor-associated factor family involved in TNF-α signalling (Wajant, et al., 2001); the complement component 5 has been associated with RA in animal models (Wang, et al., 1995). In NOAR two SNPs mapping to the *TRAF1/C5* locus (rs2900180 and rs10760130) were associated with the presence of erosions at 5 years in inflammatory polyarthritis (IP) patients; this was independent of ACPA status (Plant, et al., 2011b). At 5 years the ORs for developing erosions in IP patients after

adjusting for ACPA-positivity comprised 1.65 (95% CI 1.13-2.42; *p*=0.01) for individuals carrying the risk allele for rs2900180 and 1.52 (95% CI 1.00-2.29; *p*=0.05) for those carrying the rs10760130 risk allele. Another study of 278 cases also reported a significant association between a SNP in this locus (rs10818488) - which is in high linkage disequilibrium with the SNP identified in the NOAR cohort (rs10760130) - and radiological progression (Kurreeman, et al., 2007). A subsequent meta-analysis of 7 cohorts, containing a total of 2,666 RA patients with 6,282 radiological scores failed, however, to demonstrate an association between rs10818488 and radiological progression (*P*=0.89) (Knevel, et al., 2012a).

1.4.5.8.CD40

The CD40 protein is expressed on the surface of multiple immune cells including B cells and monocytes; it plays a pivotal role in providing helper activity by CD4+ T cells in immune reactions (Kawabe, et al., 1994). The association of the *CD40* locus was shown in 250 ACPA-positive RA cases from the Leiden EAC cohort and 393 ACPA-positive RA cases from the North American Rheumatoid Arthritis Consortium (NARAC) (Van Der Linden, et al., 2009). In this analysis the SNP, rs4810485 yielded a 1.12 (95% CI 1.04-1.21) times greater increase in the Sharp score per year in those carrying the risk genotype in the EAC. The significant association between this SNP and the rate of joint destruction was significant after correcting for multiple testing. Using a perfect SNP proxy the risk genotype from the EAC cohort also revealed a higher estimated radiological progression rate in the NARAC cohort.

1.4.5.9.IL2RA

IL2RA is an RA susceptibility locus. Its link with radiological progression in RA was demonstrated by Knevel *et al* in a meta-analysis of 1,750 RA patients across 4 independent datasets (Knevel, et al., 2013a). In this report the minor C allele of the *IL2RA* SNP, rs2104286 associated with lower radiological progression rates ($P=7.2\text{x}10^{-4}$). The pooled effect size across cohorts was 0.97 (95% CI 09.96-0.99), indicating that the C allele was associated with a 0.97-fold lower rate of joint destruction per year. The authors also provided functional evidence for this association: the minor allele associated with lower circulating levels of soluble

interleukin-2 receptor α (P=1.4x10^{-3}), which themselves associated with lower joint destruction rates (P=3.4x10^{-3}).

1.4.6. Genetics of Radiological Damage: Genome-Wide Studies

To date only two GWASs have been performed, which have evaluated genetic associations with radiological progression in RA. The first GWAS was undertaken in 384 ACPA-positive early RA patients from NARAC (Knevel, et al., 2013b). These patients had a single X-ray scored using the SvHS when they had established RA (mean disease duration 13.9 years). Genotyping was undertaken on the Illumina BeadChip (HumanHap 550k); 391,733 SNPs were available following quality control (QC) procedures. Estimated annual radiological progression rates were calculated for each patient by dividing their SvHS by the disease duration in years at the time the X-ray was performed. Linear regression with the log of the estimated annual X-ray progression rate as the response variable (SvHS are non-normally distributed) and genotype as the predictor variable was performed. P_{GWAS} was defined as $P \leq 2.7 \times 10^{-7}$. The strongest association with radiological progression was in the Sperm-Associated AntiGen 16 (*SPAG16*) locus; the relevant SNP, rs7607479 had a P-value of 1.59x10^{-7}. No other SNPs fulfilled their definition of P_{GWAS}. This finding was subsequently replicated in 3 other cohorts comprising 301 ACPA-positive RA patients from the Leiden EAC and 742 RA patients from the National Databank for Rheumatic diseases (NDB) and Witchita cohorts. In these cohorts the minor rs7607479 allele was associated with a reduced rate of radiological progression. In NARAC, patients with one minor allele had a 0.77-fold (95% CI 0.70-0.85) annual X-ray progression rate relative to the common genotype; patients carrying two minor alleles had a 0.59-fold (95% CI 0.49-0.72) annual progression rate. This SNP explained 6.6% of the variance in joint destruction between individuals in NARAC.

The function of *SPAG16* is largely unknown. To investigate its relevance to joint destruction the authors first demonstrated the presence of *SPAG16* protein in patient synovial tissue using immunohistochemical staining. This showed its expression in FLS (Knevel et al, 2013b). Secondly, they demonstrated an association between the minor rs7607479 allele and lower serum MMP-3 levels. MMPs are known to be

released by FLSs and contribute to radiological destruction in RA. They therefore provided biological evidence for a role of this variant in reducing joint destruction in RA.

The second GWAS was undertaken on the ImmunoChip (De Rooy, et al., 2013b), which although covering 195,806 SNPs and 718 small insertion-deletions associated with immune-mediated diseases, does not provide true genome-wide coverage. The ImmunoChip is an Illumina Infinium SNP microarray, which was designed in 2009 by investigators of 11 autoimmune and inflammatory disorders including RA, ankylosing spondylitis (AS) and inflammatory bowel disease (IBD). It included the top 2,000 independent association signals for each disease from meta-analyses of GWASs, alongside dense coverage of a further 186 loci within confirmed GWAS association intervals (Parkes, et al., 2013). Its goals were to enable deep replication and fine mapping of GWAS confirmed loci. This study took a different analytical approach to existing studies (De Rooy et al, 2013b). Over time genotypes can exert either a constant effect on radiological scores (at each time point the scores are different when patients are stratified by genotype; these differences remain constant over time) or an interactive effect with time on radiological scores (when patients are stratified by genotype at each time point their scores get gradually further apart). Most studies using longitudinal data have focused on this later genotype*time interaction term. The ImmunoChip study by de Rooy *et al* was undertaken in 646 early RA patients from the Leiden EAC, with 686 North American RA patients providing a replication cohort. They used a marginal regression model for longitudinal data using the log-transformed radiological score as the response variable and time, age, gender and treatment strategy as predictor variables. Two models were compared for each SNP: the first contained genotype and genotype*time interaction terms and the second contained only the clinical predictor variables. The models were compared using a likelihood ratio test (2 degrees of freedom), which tested the null hypothesis that the coefficients for both the constant SNP effect and its interaction with time equaled zero. The resultant *P*-value represented the overall genotype effect (constant and/or interactive) on radiological scores. Their *P*-value for significance was set at 1.1×10^{-6}. In the Leiden EAC 109 SNPs passed their significance threshold; 76 were in the HLA region. Of the 33 non-

HLA SNPs, 29 were available in the replication cohorts. Four SNPs were replicated; after conditional analyses 2 of these were independently associated with radiological damage: rs451066 (located on chromosome 14, downstream of the genes *ZFP36L1* and *C14orf181*) and rs11908352 (located 92 kb downstream of *CD40*-rs4810485, which is a known RA susceptibility locus). The latter variant is located within close proximity to the gene encoding for MMP-9 and fine mapping revealed it was in high LD with several variants in this region. Although these also associated with radiological severity, in a conditional analysis including these variants, only rs11908352 was significant. Subsequent serum evaluations confirmed an association between rs11908352 genotypes and serum MMP-9 levels; AA genotype carriers, which associated with more severe joint damage, had significantly higher serum MMP-9 levels compared to CC major allele carriers (P=0.007). This provided a biological mechanism through which this genotype mediates joint damage.

1.4.7. Ultrasound and MRI Imaging as Predictors of Radiological Severity

Advances in imaging technology have led to an increased use of magnetic resonance imaging (MRI) and, in particular, musculoskeletal ultrasound scanning (USS) in routine clinical practice. In early RA there is evidence that both techniques are able to predict longer-term radiological outcomes.

1.4.7.1.Ultrasound

Synovial inflammation involves periarticular vasodilation, synovial proliferation and angiogenesis; this process can be detected by the USS power Doppler (PD) modality (Jain and Samuels, 2011). USS and more specifically PD assessments have been shown to correlate with radiographic progression in several small studies. In 42 early RA patients (disease duration <12 months) followed up at 0, 3, 6 and 12 months, time-integrated values of PD USS parameters had stronger correlations with radiographic progression at 1 year (r=0.59; P<0.001) than clinical and laboratory parameters (r<0.5) (Naredo, et al., 2007). In an RCT in which 24 methotrexate treated RA cases were randomised to either placebo or infliximab, in the placebo arm there were significant positive correlations between both baseline synovial thickness and vascularity as measured by USS and progression in radiographic severity scores at 54 weeks (Taylor, et al., 2004).

1.4.7.2.MRI

Several studies have shown that the presence of MRI detected bone marrow oedema at disease onset predicts joint damage progression years later. In one RCT of 130 early RA patients baseline MRI bone marrow oedema was the only significant predictor (in a multiple linear regression analysis) of radiological progression at the wrist and MCP joints, explaining 41% of the variation in the SvHS (Hetland, et al., 2009). Similarly, in a smaller prospective study of 42 RA patients the baseline MRI bone oedema score was predictive of the 6-year total Sharp score (P=0.01) (Mcqueen, et al., 2003). Palosaari *et al* also demonstrated the predictive value of bone marrow oedema on MRI; in 27 early RA patients the baseline MRI bone oedema score was the only baseline variable that predicted erosive progression at 24 months in a multivariate model (OR 4.2; 95% CI 1.3-13.8) (Palosaari, et al., 2006).

1.4.8. Biochemical Prognostic Markers

The best established prognostic biomarkers in RA comprise the acute phase response indices, ESR and CRP, both of which correlate with disease severity (Lindqvist et al, 2005). A number of other markers have been evaluated for their prognostic implications in RA. One example is the MMPs, which are zinc dependent proteases that regulate extracellular matrix proteolysis and are involved in the cleavage of cytokines, chemokines and their receptors; they are thus considered to play important roles in inflammation (Mohammed, et al., 2003). Another example is the bone turnover marker, urinary C-telopeptide of type II collagen (CTX-II), which is an immunoassay that uses antibodies specific for the C-terminal crosslinking telopeptide of type II collagen – the most abundant protein within the cartilage matrix - in the urine (Garnero, et al., 2002a).

Young-Min *et al* evaluated the role of several serum biomarkers comprising MMP-1, MMP-13 and MMP-3, tissue inhibitor of metalloproteinases-1 (TIMP-1) and cartilage oligomeric matrix protein (COMP) and urinary biomarkers including CTX-II in predicting radiographic progression in 132 early RA patients (Young-Min, et al., 2007). They found in a multivariate analysis that a model consisting of baseline MMP-3 and CTX-II provided the best prediction of radiographic progression at study entry (AUC 0.76; 95% CI 0.66-0.85). Other research groups have shown MMP-3 to

be predictive of radiographic progression in other small RA cohorts. In 48 RA patients without radiological damage at presentation, serum MMP-3 levels at study entry significantly correlated with Sharp scores at 6 and 12 months and joint space narrowing at 6, 12 and 24 months (Posthumus, et al., 1999). Similarly, in 26 patients with early RA baseline serum MMP-3 levels were significantly associated with Larsen scores at 6 and 12 months after study entry; furthermore when the relationship between percentage increases in serum MMP-3 in the first 12 months after entry and the percentage increase in Larsen scores in each year were evaluated, a significant correlation was observed between the increase in serum MMP-3 during the first 12 months and the increase in the Larsen score in the subsequent 12–24 months after entry (Yamanaka, et al., 2000).

The role of urinary CTX-II in RA prognostic stratification has also been reproduced in several studies. The association between baseline urinary CTX-I and CTX-II levels and the mean annual progression of joint destruction over a median of 4 years was examined in the COBRA study. In two multivariate logistic regression analyses that included each marker separately due to their high correlation, baseline urinary CTX-I and CTX-II levels both predicted long-term radiologic progression independently of treatment, disease activity and RF status at baseline (Garnero, et al., 2002b). Additionally, Hashimoto *et al* reported that in 145 patients with active RA of less than 5 years duration baseline urinary CTX-II levels correlated significantly with radiological progression at week 52 (Hashimoto, et al., 2009).

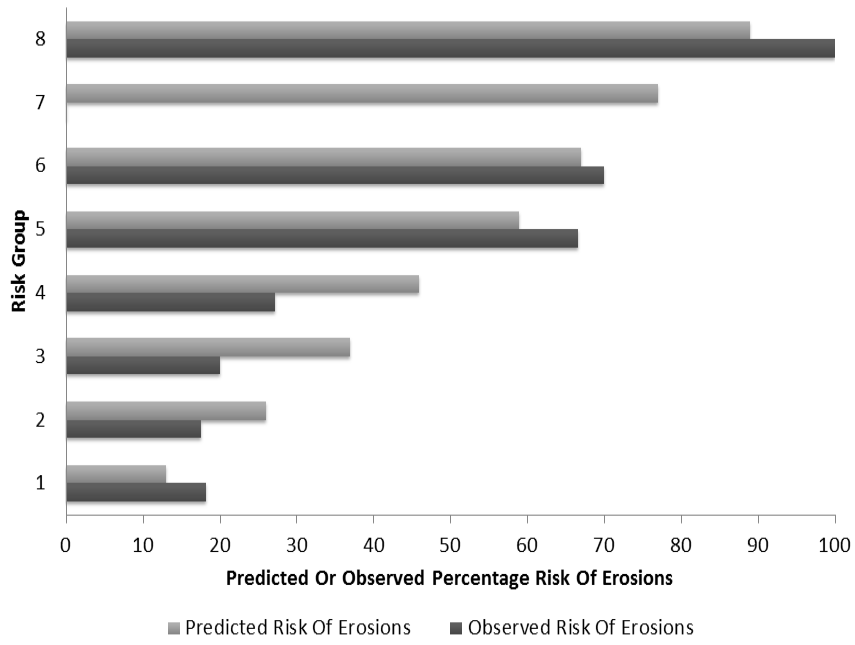
1.4.9. Prognostic Modelling In RA

Several research groups have attempted to combine prognostic factors into models capable of identifying individuals at a high-risk of radiological progression. Some have used simple clinical parameters; others have integrated these with biomarkers and radiological indices. Genetic markers have rarely been used.

Brennan *et al* developed a prediction model for the development of erosions in the hands and/or feet after 12 months in NOAR (Brennan, et al., 1996). In this study of 175 patients with early RA, the population was randomly split into a prediction sample of 105 patients, in whom predictor variables for radiological progression

were sought, and a validation sample of 70 patients, in whom the prediction algorithm was tested. A simple algorithm using a combination of three variables (positive RF test, swelling in ≥2 large joints and disease duration >3 months) was best able to predict erosions. This model was able to classify 8 risk groups with a probability of developing erosions ranging from 0.13 (all variables absent) to 0.89 (all variables present). It correctly predicted the development of erosive disease in 79% of cases. The model's negative and positive predictive values were 80% and 76%, respectively. Its predictive abilities are illustrated in Figure 1-9.

Figure 1-9. Predicted Vs. Observed Risks of Erosions Using a Prediction Model with Three Clinical Variables in NOAR



Risk groups defined by the presence/absence of 3 variables; combination of variables comprising a positive rheumatoid factor, disease duration ≥ 3 months and ≥ 2 large joints involved in group 1 = negative/no/no; group 2 = negative/yes/no; group 3 = negative/no/yes; group 4 = positive/no/no; group 5 = negative/yes/yes; group 6 = positive/yes/no; group 7 = positive/no/yes; group 8 = positive/yes/yes. Figure adapted using data from Brennan et al (Brennan et al, 1996) and reproduced with permission from Scott et al (Scott et al, 2013a).

Drossaers-Bakker *et al* demonstrated that prognostic modelling can be undertaken to predict longer-term disease outcomes at 12 years (Drossaers-Bakker, et al., 2002).

This study evaluated 112 female RA patients with symptoms of less than 5 years duration (median 1 year) at recruitment. It developed prediction models for 3 different disease outcomes: firstly radiographic damage (measured by the SvHS method), secondly disability (measured by the HAQ) and thirdly a severe disease course (measured by calculating the area under the curve of all DAS assessments alongside the radiographic disease course). Individuals in the highest tertile of each outcome measure were defined as "severe" for that outcome and individuals in the lowest tertile were defined as "mild". Using a model that contained the baseline parameters the SJC, RF, the presence of erosions, the Ritchie index, ESR, HAQ and SvHS the accuracy of the model for predicting mild radiographic damage, severe radiographic damage, mild HAQ, severe HAQ and a severe disease course comprised 87%, 84%, 88%, 84% and 83% respectively. Surprisingly, additional information on HLA typing added little to the modelling, improving the correct prediction of radiographic damage by only 3%. This study, however, developed and validated its model within the same patient cohort; it is therefore expected that their model was able to define disease outcomes with relative accuracy. It requires external validation.

More recently two research groups have developed matrix risk models for rapid radiological progression (RRP), which are organised into colour coded matrixes similar to that which is widely used in predicting the 10-year risk of fatal cardiovascular disease (Vastesaeger, et al., 2009). One of these matrixes was developed using data from 465 RA patients enrolled to the BeST RCT. This study randomised patients to four treatment arms comprising two arms treated with initial monotherapy that could be switched or extended to other DMARDs, a third arm treated with initial combination DMARDs and tapering high dose corticosteroids and a fourth arm treated with initial methotrexate and infliximab (Visser, et al., 2010). Patients were treated with an aim of attaining a DAS of ≤2.4. RRP was defined as an increase in the SvHS of ≥5 after 12 months. Predictors of RRP were identified by multivariate logistic regression with backward selection. Different models were developed for different treatment groups and included the variables CRP, erosion score and serology (RF and ACPA). The highest risk group was those individuals in the initial monotherapy treatment arm with a CRP ≥35mg/L, erosion score ≥4 and

both RF and ACPA positivity; their risk of RRP was 78%. The lowest risk groups were those individuals in the initial combination treatment with prednisolone or infliximab arms with a CRP <10mg/L, erosion score of 0 and negative serology; their risk of RRP was 1%. The AUC was 0.81 (95% CI 0.77-0.86) indicating a moderate ability to correctly classify individuals who will develop RRP. This study, however, also validated its prediction model within the prognostic factor identification cohort; this model therefore requires evaluating in another patient cohort to better define its prognostic capabilities.

The other research group to develop a risk matrix for RRP developed prediction models in two different cohorts (Vastesaeger et al, 2009). The first cohort was the ASPIRE study, which comprised 1,049 methotrexate-naive early RA patients randomised to receive methotrexate with or without infliximab; the second cohort was the ATTRACT trial, which comprised 428 patients with established RA and active disease treated with either methotrexate and infliximab or placebo. They identified risk factors from the early RA cohort (the ASPIRE study) and in order to ensure this combination of risk factors had similar predictive capabilities in a more advanced RA population undergoing similar treatment generated a prediction model in the ATTRACT trial using the same variables. RRP was defined as a change in the modified SvHS of ≥5 units/year. Spearman's rank analysis was used to identify baseline risk factors for RRP. Two prediction matrixes were developed, which contained either the ESR or CRP alongside information on the 28 SJC, RF and treatment (monotherapy or combination therapy). The highest risk group was those individuals in the ATTRACT study receiving methotrexate monotherapy with a 28 SJC >17, RF titre >200U/ml and an ESR >50mm/h; their risk of RRP was 65%. The lowest risk group was those individuals in the ASPIRE study receiving methotrexate and infliximab with a 28 SJC <10, RF <80U/ml and an ESR <21mm/h; their risk of RRP was 2%. Individuals treated with methotrexate monotherapy had higher predicted rates of RRP when compared to those receiving infliximab.

The main criticism of these models is that, with the exception of the latter model, they have not been validated in external populations. Additionally, with regards to the last two matrix models they were developed within a clinical trial setting; these

patients usually have more aggressive disease and receive more rigorous treatment than they would in a routine clinical setting. It is, therefore, unclear if their results are generalisable to more typical RA populations.

1.5. Predicting Treatment Responses in RA

1.5.1. Why Treatment Response Predictors Are Needed

Current NICE guidelines advocate a single approach to the management of RA patients. In individuals with early active disease, combination DMARDs and short-term corticosteroids are recommended with an escalation to biologics in refractory cases (Deighton et al, 2009). These guidelines provide no advice on which specific agents (beyond methotrexate as an anchor DMARD) should be used. This reflects uncertainties around which factors predict an individual's likely response to a specific drug. As a result, patients are often exposed to trials of treatments that may not benefit them. This has important personal implications, potentially exposing patients to the side-effects of ineffective treatments, alongside the associated complications of sustained disease activity whilst an effective drug is found through trial and error. It also has societal implications, with the high-costs of ineffective trials of biologics in individual patients using health care funding that may be better spent in other areas of the health service. Identifying predictors of treatment responses is therefore a key research goal.

To a certain extent predictors of RA severity also overlap with predictors of treatment responses. Risk factors that identify patients with a severe disease course indicate they are less likely to respond to non-intensive treatment regimens. One example of this is ACPA status, with a secondary analysis of the BeST study indicating that methotrexate monotherapy is unlikely to provide adequate disease control in ACPA-positive patients (De Vries-Bouwstra, et al., 2008). There are, however, a number of factors that associate with the efficacy of specific therapeutic agents; these are outlined below.

1.5.2. Predictors of Synthetic DMARD and Biologic Agent Efficacy

A number of factors have been suggested to predict treatment responses to synthetic DMARDs and biologics. Overall, the results are inconsistent. In a systematic review of DMARD response predictors, Hider *et al* attributed conflicting results to methodological heterogeneity across studies (Hider, et al., 2005). Examples of this heterogeneity included the use of variable DMARD response definitions, the evaluation of different DMARD agents and different DMARD regimens (some studies assessed monotherapy and others combination therapy). Such heterogeneity makes drawing definitive conclusions on treatment response predictors problematic.

1.5.2.1.Gender

Overall, the evidence suggests that females with RA probably respond more poorly to methotrexate than males. A systematic review by Drouin *et al* reported that female gender associated with a poor methotrexate response in all 9 studies included in their review; the reported ORs for a clinical response in women vs. men ranged from 0.3 to 0.7 (Drouin, et al., 2010). Similarly, in an analysis of the SWEFOT trial female gender was associated with a reduced likelihood of a EULAR response in 487 early RA patients treated with methotrexate monotherapy (OR 0.50; 95% CI 0.31-0.81) (Saevarsdottir, et al., 2011a). Other studies have replicated these findings (Stranzl, et al., 2003).

A gender effect on treatment responses has not been demonstrated with other DMARDs. Capell *et al* found no effect of gender on responses to gold, penicillamine, sulfasalazine, or auranofin in 1,140 RA patients (Capell, et al., 1993). Similarly, no effect of gender on responses to gold, sulfasalazine, or penicillamine was found in an observational study of 681 RA patients (Situnayake and Mcconkey, 1990).

There is some evidence to suggest that females may respond less favourably to anti-TNF. An analysis of 2,879 patients from the British Society for Rheumatology Biologics Register (BSRBR) found that female gender associated with a reduced odds of remission with etanercept (OR 0.61; 95% CI 0.38–0.94) and infliximab (OR 0.60; 95% CI 0.40–0.89) treatment (Hyrich, et al., 2006). In another analysis of 1,257

established RA patients an association was found between male gender and an increased likelihood of attaining remission after 6 months of anti-TNF treatment (Mancarella, et al., 2007). This gender effect has, however, not been reproduced in all populations with an analysis of the South Swedish Arthritis Treatment Groups Register reporting no gender effect on anti-TNF responses (Kristensen, et al., 2008).

1.5.2.2.Ethnicity

There is no definitive evidence that ethnicity influences treatment responses. One retrospective analysis of 1,191 North European and 193 South Asian patients found that individuals in the latter ethnic group terminated DMARD therapy significantly earlier (Helliwell and Ibrahim, 2003). The 12-month survival rates of DMARDs in North Europeans and South Asians were 0.74 (95% CI 0.69-0.79) and 0.67 (95% CI 0.65-0.68), respectively (P<0.00001). Although inefficacy was one reason for discontinuation, there were other potential reasons including side-effects. In another prospective study of 134 patients with early RA that were DMARD-naive at inception no relationship was seen between ethnicity and DMARD response at 12 months (defined as the attainment of low disease activity assessed using the Simplified Disease Activity Index) (Hodkinson, et al., 2012). This study was limited by its small sample size, short follow-up period and unrestricted DMARD use. There is a lack of data on whether ethnicity affects responses to biologics.

1.5.2.3.Age

Age does not appear to affect methotrexate responses. A meta-analysis of 11 clinical trials evaluating methotrexate reported no association between age and treatment efficacy in 496 RA patients (Rheumatoid Arthritis Clinical Trial Archive Group, 1995). One observational study found a significant association between a younger age and a positive response to sulfasalazine; no age effect was seen on responses to gold or penicillamine (Situnayake and Mcconkey, 1990).

Although data from the BSRBR suggests that age does not impact on anti-TNF efficacy (Hyrich et al, 2006), there is some evidence that it may influence tocilizumab response. In an observational study of 204 RA patients in France, a younger age (<55 years) positively correlated with EULAR response and remission

rates at 6 months in patients treated with tocilizumab (Pers, et al., 2014). Data from the REACTION study supported this finding, with a younger age associating with an increased chance of remission in 229 RA patients treated with tocilizumab (Yamanaka, et al., 2011).

1.5.2.4.Smoking

It appears likely that current smokers respond less favourably to methotrexate and anti-TNF. This relationship has been reported in a number of large observational studies. An analysis of the EIRA study revealed that current smokers were less likely to achieve a good response at 3 months after initiating methotrexate (P=0.05) or anti-TNF (P=0.03) (Saevarsdottir, et al., 2011b). In multivariate analyses adjusting for other relevant clinical, serological, and genetic factors, the inverse associations between current smoking and good treatment responses remained; the adjusted ORs for methotrexate and anti-TNF response in current vs. never smokers were 0.60 (95% CI 0.39-0.94) and 0.52 (95% CI 0.29-0.96), respectively. These differences persisted at later follow-up visits. Interestingly, past smoking did not affect the chance of a treatment response. The association between current smoking and poor methotrexate response was reproduced in the SWEFOT trial (Saevarsdottir et al, 2011a). In this study current smoking significantly reducing the odds of a EULAR response after 3-4 months of treatment with methotrexate; the adjusted OR was 0.35 (95% CI 0.20-0.63). Registries also indicate that current smokers are less likely to respond to anti-TNF with data from the BSRBR demonstrating that current smoking associates with a reduced likelihood of etanercept and infliximab response (Hyrich et al, 2006).

1.5.2.5.Serology

In a recent comprehensive literature review, Romão *et al* highlighted a lack of robust evidence that RF status predicts DMARD responses (Romao, et al., 2013). Similarly, ACPA does not seem to associate with individual DMARD responses. Two studies of early RA found no association between ACPA and methotrexate response (Wessels, et al., 2007, Saevarsdottir et al, 2011a). Gosec *et al* also reported no influence of ACPA on remission rates in 191 early RA patients receiving a range of different DMARDs over 5 years of follow-up (Gossec, et al., 2004).

Serological status may affect biologic responses. Clinical trial data suggests that rituximab is slightly more effective in seropositive disease. A meta-analysis of four RCTs identified a modest but significant effect of seropositivity (presence of RF and/or ACPA) on the response to rituximab; the 24 week reduction in DAS28 scores in seropositive patients was 0.35 units (95% CI 0.12 to 0.84) larger than in seronegative patients (Isaacs et al, 2013). Abatacept also appears to be more effective in ACPA-positive RA with the Orencia Rheumatoid Arthritis (ORA) registry reporting that ACPA positivity associated with good EULAR responses in RA patients receiving abatacept (76% and 62% of good/moderate and non-EULAR responders, respectively were ACPA-positive) (Gottenberg et al, 2012). By contrast TNF-inhibitors seem more efficacious in ACPA-negative disease. In an analysis of 617 RA patients from the Rheumatic Diseases Portuguese Register, ACPA provided an OR of 0.97 (95% CI 0.95-0.98; $P<0.0001$) for attaining a EULAR good response when treated with adalimumab, etanercept or infliximab. Potter *et al* also reported that in 642 UK RA patients, those that were ACPA-negative had a 0.39 (95% CI 0.07 to 0.71) greater mean improvement in DAS28 when treated with anti-TNF compared to patients that were ACPA-positive (Potter et al, 2009).

1.5.2.6.Disease Duration

It is well established that very early treatment in RA associates with better outcomes (Breedveld, 2011). This concept, termed the "window of opportunity" may be as short as the first 3 months of disease (Raza, et al., 2012). This is demonstrated in a meta-analysis of 14 RCTs, which evaluated predictors of treatment responses in a variety of DMARD regimens. Disease duration was a major predictor of treatment responses (Anderson, et al., 2000). In a pooled analysis among patients receiving active treatment, the percentage of patients attaining an ACR 20 response comprised 53%, 43%, 44%, 38% and 35% for patients with disease durations of <1 year, 1- 2 years, 2–5 years, 5–10 years and >10 years, respectively.

There does not appear to be an effect of disease duration on anti-TNF responses. The BSRBR reported no association between disease duration and responses to anti-TNF therapy, although most patients in their cohort had established disease (mean disease duration 14 years) (Hyrich et al, 2006). Similarly, data from the Swedish registry

indicated no effect of disease duration on anti-TNF responses; again most individuals had established disease with a mean duration of 11 years in males and 12 years in females (Kristensen et al, 2008).

1.5.2.7.Genetics

Multiple studies have attempted to identify genetic loci in the methotrexate cellular pathway that predict methotrexate responses in RA. Their sample sizes are small and their findings often not replicated. A recent systematic literature review of this topic identified 4 methotrexate cellular pathway SNPs that were examined in 5 or more studies for their association with methotrexate efficacy and toxicity (Malik and Ranganathan, 2013). These comprised two SNPs in the *MTHFR* gene (rs1801133 and rs1801131), one in the *SLC19A1* gene (rs1051266) and another in the *ATIC* gene (rs2372536). Of the 20 studies evaluating the association between rs1801133 and methotrexate efficacy only 3 reported a significant result; of these 1 reported greater efficacy in patients carrying the T allele of this SNP and the other 2 reported lower efficacy in patients carrying the T allele. Similar inconsistencies were seen in studies evaluating the other 3 SNPs. All studies have used a candidate gene approach; no genome-wide studies have evaluated genetic loci associated with methotrexate response.

To date three GWASs have been undertaken to evaluate genetic associations with anti-TNF responses. Two of these identified several loci at a level of significance suggestive of an association (Plant, et al., 2011a, Liu, et al., 2008). They were, however, of limited sample sizes (one comprised 89 and the other 566 patients) and therefore had limited power to detect SNPs attaining P_{GWAS}. The third GWAS evaluated a substantially larger patient cohort, comprising 2,706 RA patients from 13 collections (Cui, et al., 2013). All patients had received etanercept, infliximab or adalimumab. DAS28 scores were collected at baseline and at one time point after anti-TNF therapy administration. This study evaluated genetic associations with the change in DAS28 score from baseline. One SNP (rs6427528) at the *1q23* locus associated with change in DAS28 in the etanercept subset of patients ($P=8\times10^{-8}$). This SNP was predicted to disrupt transcription factor binding site motifs in the 3′ UTR of an immune-related gene, *CD84*. The allele that associated with a better

etanercept response also significantly associated with higher *CD84* gene expression in peripheral blood mononuclear cells.

A recent GWAS has also evaluated predictors of tocilizumab response in 1,683 RA patients from 6 clinical studies (Wang, et al., 2013). To optimise the power to detect relevant associations a conservative definition for a likely association with changes in RA outcomes (DAS28, SJC, TJC, HAQ, CRP) and ACR20 responses was used. The authors selected candidate SNPs for replication that met one or more of the following criteria: (1) $P<10^{-5}$ in white subjects; (2) $P<10^{-4}$ in White subjects and a lower *P*-value in all ethnicities; or (3) those selected by least absolute shrinkage and selection operator (LASSO) penalized regression analysis of White subjects or all ethnicities. Using this definition 207 markers were identified as having 253 significant associations. Seven of these achieved confirmation at $P<0.05$ with the same outcome in the replication cohort. None of these markers were obviously linked to the IL-6 pathway. In total 4 SNPs attained true P_{GWAS}; none of these were replicated. The study authors concluded that it is unlikely that a major genetic determinant of tocilizumab response exists.

1.5.2.8.Other Biomarkers

A number of studies have examined other biomarkers for their impact on treatment responses. Although promising these markers lack clinical utility, as they have generally been examined in small sample sizes, have not been replicated in larger patient cohorts or associate with only small differences in treatment responses. Examples include serum MMP-3 levels (Posthumus, et al., 2002) and red blood cell levels of methotrexate polyglutamates (Dervieux, et al., 2005), which have been studied for their association with methotrexate responses and stimulated whole blood cell pro-inflammatory cytokine levels (Kayakabe, et al., 2012) and serum proteins (Ortea, et al., 2012), which have been studied for their association with anti-TNF efficacy.

1.5.3. Prediction Models for Treatment Responses in RA

To a certain extent the previously described matrix models that predict the risk of RRP with different RA treatment regimens can be considered to predict responses to

those treatment strategies (Vastesaeger et al, 2009, Visser et al, 2010). Another research group has developed a model that predicts response to a specific therapeutic agent (methotrexate) within 205 early, active RA patients (from the methotrexate monotherapy treatment arm of the BeST study) (Wessels et al, 2007). Four clinical and four genetic variables were used to generate a clinical score, which predicted the likelihood of a treatment response (defined as a DAS ≤2.4 at 6 months). These comprised sex, RF, smoking status, baseline DAS, and 4 SNPs (in *AMPD1*, *ATIC*, *ITPA*, and *MTHFD1* loci). There variables were chosen from a panel of 24 potential baseline variables through backward selection. The clinical score used simplified regression coefficients of the independent variables, providing a score ranging from 0 to 11.5. Using a score cut-off of ≤3.5 points for responders provided a true positive rate of 95%; using a score cut-off of ≥6 points for non-responders provided a true negative response rate of 86%. The R^2 value for the model was 0.35; the AUC following cross-validation was 0.79.

This model has been externally validated in two small replication cohorts. The first cohort comprised 38 early RA patients (Wessels et al, 2007); the true positive response rate was 70% and the true negative response rate was 72%. No AUC values were provided for this replication cohort. The second cohort comprised 75 patients with established RA (Fransen, et al., 2012); the true positive rate was 47% and the true negative response rate was 81%. The AUC value in this cohort was 0.77 and the R^2 0.28. Therefore, despite similar levels of discrimination between responders and non-responders and similar abilities to predict non-responders, the model was substantially less accurate at predicting methotrexate responders in established RA, when compared with the early RA cohort. An important limitation of this model is the size of its intermediate risk group (approximately half of actual responders are classified as having an "intermediate risk" of responding). Identifying risk factors that account for the remaining variance in methotrexate response should reduce the size of this group. In addition the model requires validation in larger sized, prospective RA cohorts. Similar prediction models are yet to be developed for other DMARDs or biologics.

1.6. Aims and Objectives

1.6.1. Research Rationale

RA is a highly heterogeneous disorder. Its variability can be seen from several perspectives. Firstly, there are substantial differences in the risk factors which precipitate RA in different patients. The two dominant RA susceptibility factors – cigarette smoking and *HLA-DRB1* SE allele carriage – are not present in all RA patients. In addition not all smokers with the SE develop RA. Secondly, there is marked variation in the clinical course of RA. Some patients have an aggressive phenotype, which is characterised by the early development of radiological erosions and disability. In other patients there is a relatively benign course with normal radiographs and no functional decline. Finally, RA patients vary in their responses to treatment. An important example of this variability is that one third of RA patients fail to achieve ACR 20 responses after being treated with a TNF-inhibitor (Rubbert-Roth and Finckh, 2009), which is the standard treatment advocated for all patients with active, DMARD-refractory disease.

One important consequence of the heterogeneity of RA is that delivering high quality care depends upon adopting a stratified approach to its management (a concept termed "stratified medicine"). This stratified approach uses clinical characteristics and biomarkers to identify groups of individuals that are most likely to respond to specific treatment strategies (Plant, et al., 2014). Such strategies include not only the treatment of established disease but also its prevention in high-risk patient strata. This approach contrasts with current NICE RA guidelines, which advocate empirical practice with patients managed as a single group (Emery et al, 2010).

Stratified medicine and risk prediction are overlapping concepts. Stratified medicine involves identifying subgroups of patients with distinct mechanisms of disease or particular treatment responses (Medical Research Council, 2014). Such an approach enables the identification and development of treatment strategies, which are effective for specific groups of patients. Its ultimate goal is to ensure the right patient gets the right treatment at the right time (Medical Research Council, 2014). Risk prediction involves developing models in which one or more predictors are used to estimate the risk that one or more outcomes are present (diagnostic prediction model)

or will occur within a specific time period (prognostic prediction model) in an individual with a particular predictor profile (Moons, et al., 2012). Risk prediction therefore facilitates stratified medicine. Identifying relevant predictive factors and incorporating them in accurate risk prediction models is a crucial first step in moving from empirical clinical practice towards following the ethos of stratified medicine in clinical care.

1.6.2. Overall Aim

For risk prediction and stratified medicine in RA to be possible the following two requirements must be met. Firstly, the factors that define subgroups of individuals likely to a) develop RA (and thus benefit from preventive treatments); b) have a severe disease course requiring aggressive treatment; and c) respond to specific medications must be identified. Secondly, prediction modelling frameworks are required that harness these factors to stratify individuals into subgroups that are either likely to develop RA or have an established RA phenotype that is likely to benefit from a specific management strategy.

This thesis addresses these two requirements. Its overall goal is to facilitate the risk prediction that underpins stratified medicine in RA. Its primary aim is to improve the knowledge of which clinical and genetic factors predict RA's onset, disease course and treatment responses. Its secondary aim is to develop a risk prediction modelling framework that harnesses these factors to inform clinical care. Due to this broad conceptual framework, the thesis does not test a single hypothesis; instead it tests a series of five inter-related hypotheses, which are outlined below.

1.6.3. Specific Objectives

The research in this thesis addresses the overall aim by focusing on five specific objectives. The first objective is to evaluate published data on a proposed environmental risk factor for RA (alcohol abstinence) to improve the understanding of its relevance to disease development. The second and third objectives are to identify novel genetic predictors for RA severity using candidate gene and genome-wide approaches, respectively. The fourth objective is to evaluate the role of serology (ACPA status) in predicting treatment needs and responses in early, active

RA patients. The final objective is to incorporate published RA risk factor data within a novel prediction modelling framework to better understand how such factors can be applied to stratify individuals into disease-risk groups.

1.6.3.1.Evaluating Alcohol as a Protective Factor against RA

Recent studies have reported an inverse association between alcohol consumption and the risk of RA development. This suggests that alcohol may protect against RA. As with many environmental factors this relationship is mainly observed in case-control (Kallberg et al, 2009, Maxwell et al, 2010) and not cohort studies (Cerhan et al, 2002); a causative role is therefore uncertain.

This research objective is to establish if alcohol intake protects against RA development and to determine if this effect is influenced by alcohol dose, duration and serological status. It tests the hypothesis that alcohol intake influences the likelihood of RA development and that this association varies by serological status and the level and duration of alcohol intake. The research addressing this objective is presented in the form of a published manuscript (Scott et al, 2013c) in Chapter 2 of this thesis.

1.6.3.2.Genetic Predictors of X-ray Progression: Candidate Gene Approach

Radiological damage is considered an important marker of RA severity with the presence of X-ray damage often used to guide treatment decisions in clinical practice (Garrood, et al., 2011). Identifying patient-specific factors that predict radiological damage are, therefore, highly desirable. Despite the moderate heritability of RA radiological progression rates (Knevel et al, 2012b) there is limited data on the genetic variants influencing this trait.

Several studies have tested the hypothesis that RA susceptibility variants associate with radiological progression. Their findings are often not replicated; additionally they have examined only a proportion of risk loci and evaluated each locus separately, despite limited power. Two recent meta-analyses of genetic studies have expanded the number of genetic loci and HLA protein amino acid polymorphisms associated with RA susceptibility (Okada et al, 2013, Raychaudhuri et al, 2012). This

research objective is to establish if these variants associate with radiological progression in early, active RA patients, when they are assessed both individually and cumulatively, using a genetic risk score combining risk loci. It tests the hypothesis that RA genetic susceptibility variants associate with radiological progression in patients with early, active disease. The research addressing this objective is presented in the form of a traditional thesis section in Chapter 3.

1.6.3.3.Genetic Predictors of X-ray Progression: Genome-Wide Approach

This objective is directly related to objective 3. As opposed to taking a candidate gene approach (examining the role of susceptibility variants in radiological progression) it will attempt to identify novel genetic associations with X-ray progression on a genome-wide scale (covering 138,488 genetic markers with established links to immune-mediated diseases present on the ImmunoChip). This research objective is to identify novel genetic associations with radiological progression in early, active RA patients. It tests the hypothesis that genetic markers present on the ImmunoChip associate with radiological progression in patients with early, active RA. The research addressing this objective is presented in the form of a traditional thesis section in Chapter 4.

1.6.3.4.ACPA as a Predictor of Treatment Needs and Responses

Current UK guidelines recommend that all patients with early, active RA receive combination DMARDs and short-term corticosteroids (Deighton et al, 2009). It is uncertain if this approach is relevant to all patients. Evidence suggests that DMARD monotherapy may be sufficient in some individuals (De Vries-Bouwstra et al, 2008). Additionally, studies indicate that differential treatment responses exist between ACPA RA subsets (Isaacs et al, 2013, Potter et al, 2009). When this is considered alongside the fact that ACPA-positive and ACPA-negative RA have different susceptibility factors and disease courses (Daha and Toes, 2011), it appears that ACPA status may be an important biomarker in establishing patient subgroups that are likely to respond to specific treatments. This research objective is to establish if patients with early, active ACPA-positive and ACPA-negative RA differ in their responses to intensive treatment with combination DMARDs and corticosteroids. It tests the hypothesis that responses to intensive combination treatments in early,

active RA patients differ by ACPA status. The research addressing this objective is presented in the form of a published manuscript (Seegobin, et al., 2014) in Chapter 5 of this thesis.

1.6.3.5.Developing and Validating a Risk Prediction Model for RA

In order for stratified medicine to be possible, risk prediction modelling frameworks are required that use information on patient-specific factors to stratify individuals into subgroups likely to benefit from specific management strategies. As outlined in the introductory section of this thesis, the factors that predict an RA patient's likely disease course and treatment response are poorly defined. In contrast, the genetic factors that predict an individual's risk of RA development are well established, with the most recent meta-analysis of GWASs identifying over 100 RA genetic susceptibility loci (Okada et al, 2013). Developing a prediction modelling framework that uses these genetic factors to stratify patients to disease-risk groups is of crucial importance. Such a model could be used to identify individuals at a very high-risk of RA with a longer term goal of evaluating primary prevention strategies. Furthermore, once validated, such a framework could be applied to many aspects of stratified medicine, combining patient-specific factors for disease severity and treatment responses (once they are established) to inform management decisions in an established RA phenotype.

Previous attempts to combine RA genetic susceptibility factors in risk prediction models have failed to generate clinically relevant predictive data (Chibnik et al, 2011, Karlson et al, 2010, Karlson et al, 2013, Yarwood et al, 2013). This research objective is, therefore, to evaluate if an alternative approach to prediction modelling (using computer simulation to categorise risk) will generate clinically applicable data and to establish if it can better identify younger onset RA (YORA), an RA subset that is particularly desirable to prevent due to higher associated health-care costs (Lajas, et al., 2003). It tests the hypothesis that RA risk factors can be combined to predict an individual's future risk of disease development. The research addressing this objective is presented in the form of a published manuscript (Scott, et al., 2013b) in Chapter 6 of this thesis.

CHAPTER 2. ALCOHOL AS A PROTECTIVE FACTOR

This chapter is presented as a published paper and is a copy of the following journal publication:

Scott IC, Tan R, Stahl D, Steer S, Lewis CM, Cope AP. The protective effect of alcohol on developing rheumatoid arthritis: a systematic review and meta-analysis. Rheumatology (Oxford) 2013; 52: 856-67.

This publication is available at:
http://rheumatology.oxfordjournals.org/content/52/5/856.long.

RHEUMATOLOGY

Rheumatology 2013;52:856–867
doi:10.1093/rheumatology/kes376
Advance Access publication 3 January 2013

Original article

The protective effect of alcohol on developing rheumatoid arthritis: a systematic review and meta-analysis

Ian C. Scott[1,2], Rachael Tan[1], Daniel Stahl[3], Sophia Steer[4], Cathryn M. Lewis[2] and Andrew P. Cope[1]

Abstract

Objectives. Our aim was to establish whether alcohol protects against RA development and to determine whether this effect is influenced by alcohol dose, duration and serological status through systematically reviewing the literature and undertaking a meta-analysis.

Methods. We searched Medline/EMBASE (1946 to July 2012) using the terms rheumatoid arthritis.mp or arthritis, rheumatoid/ and alcohol.mp or ethanol/. Manuscript bibliographies were reviewed. Observational studies were included that were case–control/cohort, examined the relationship between alcohol and RA risk and reported or allowed the calculation of effect size data [odds ratios (ORs)/relative risks (RRs) with 95% CIs] in drinkers *vs* non-drinkers. A random-effects model was used to estimate pooled ORs/RRs. Dose–risk relationships were evaluated by trend tests.

Results. Nine studies (from 893 articles) met our inclusion criteria, comprising six case–control (3564 cases; 8477 controls) and three cohort studies (444 RA cases; 84 421 individuals). A significant protective effect of alcohol on RA risk was observed—summary OR for RA in drinkers *vs* non-drinkers 0.78 (95% CI 0.63, 0.96). This effect was confined to ACPA-positive RA—summary OR 0.52 (95% CI 0.36, 0.76), with no significant risk reduction seen for ACPA-negative RA—summary OR 0.74 (95% CI 0.53, 1.05). Subgroup analysis by study design identified a significant relationship in case–control but not cohort studies.

Conclusion. Alcohol intake is inversely associated with ACPA-positive RA, suggesting a protective effect. As this finding is confined to case–control studies further research is required with prospective cohort studies incorporating ACPA status to confirm this relationship.

Key words: rheumatoid arthritis, alcohol, cyclic peptides, systematic review.

CLINICAL SCIENCE

Introduction

Many risk factors have been implicated in the development of RA [1]. Their impact varies according to patients' RF and antibodies to citrullinated protein antigen (ACPA) status [2]. Genetic risk factors have been studied in the most detail with over 30 risk loci established for seropositive RA [3]. Smoking is the main environmental risk factor identified to date, which also predominantly predisposes to seropositive RA [4]. The roles of other environmental risk factors are less clear. The emergence of risk prediction models, which combine gene–environment factors to identify individuals at a high risk of RA, highlight the importance of accurately defining RA's underlying risk factors [5].

Recent case–control studies show that fewer RA patients drink alcohol when compared with controls; this finding suggests that alcohol intake may protect against the development of RA [6, 7]. As with smoking this relationship is greater for ACPA-positive RA and increases with exposure. This beneficial effect of alcohol, which has attracted substantial media interest, has not been

[1]Academic Department of Rheumatology, Centre for Molecular and Cellular Biology of Inflammation, [2]Department of Medical and Molecular Genetics, Guy's Hospital, King's College London, Great Maze Pond, [3]Department of Biostatistics, Institute of Psychiatry, De Crespigny Park and [4]Department of Rheumatology, Weston Education Centre, King's College Hospital, London, UK.

Submitted 18 July 2012; revised version accepted 2 November 2012.

Correspondence to: Ian C. Scott, Department of Medical and Molecular Genetics, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.
E-mail: ian.scott@kcl.ac.uk

identified in a number of earlier studies [8, 9]. As a consequence its significance is controversial.

We therefore systematically reviewed observational studies evaluating the relationship between alcohol intake and the development of RA and undertook a meta-analysis. Our primary aim was to examine if alcohol affected the risk of RA, through testing the hypothesis that alcohol intake influenced the likelihood of RA development. Our secondary aims were to establish if this relationship varied by serological status and according to the level and duration of alcohol consumed, through testing the hypotheses that alcohol intake predominantly affected seropositive RA risk and that this risk varied according to the level and duration of alcohol intake.

Methods

Reporting structure and data extraction

We adopted the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) checklist for the reporting of this review [10]. Our literature search, data extraction and study quality assessments were performed in an independent, unblinded manner by two authors (I.C.S. and R.T.). Discrepancies were resolved by discussion. Inclusion criteria and analytical methods were pre-specified in a protocol.

Search strategy

Medline (from 1946 to July 2012) and EMBASE (from 1947 to July 2012) were searched using the Ovid platform (search last performed July 2012). The search terms comprised: rheumatoid arthritis.mp or arthritis, rheumatoid/ and alcohol.mp or ethanol/. Reference lists of included manuscripts were reviewed for relevant papers. EndNote version X5 (Thomson Reuters, NY, USA) was used for citation management.

Selection criteria

Observational studies were included that: (i) were case–control or cohort design; (ii) examined the relationship between alcohol intake and the risk of RA development; and (iii) reported effect size data as odds ratios (ORs) or relative risks (RRs) with 95% CIs or provided data allowing for their calculation.

We excluded: (i) additional studies evaluating the same cohort (only the largest study with the most complete information was included); and (ii) unpublished studies (conference abstracts). All identified abstracts were published in English; we did not identify any studies in non-English languages fulfilling our inclusion criteria from abstract review.

Data extraction

We extracted the following data: author and manuscript names, study design, publication year, sample size, demographics (age and gender), RA characteristics (disease duration, disease activity, serology and radiographic erosions), information on alcohol intake, effect size data (ORs or RRs with 95% CIs), adjustment factors included in the analyses and geographical area.

Assessment of study quality

To evaluate the validity of the included studies their quality was assessed using the Newcastle–Ottawa Scale (advocated by the Cochrane Non-Randomised Studies Methods Working Group) [11, 12]. This provides points (termed stars) for eight items across three domains comprising group selection (maximum 4 points), comparability of cohorts or cases and controls (maximum 2 points) and ascertainment of outcome of interest or exposure (maximum 3 points). Individual components of each domain are detailed in supplementary Tables S1 and S2 (available as supplementary data at *Rheumatology* Online). The total score ranges from 0 to 9. Study quality scores have been criticized within the literature because they often incorporate diverse items that are weighted differently [13]; we have, therefore, provided a breakdown score for each domain subscale in addition to the total summary score.

Statistical analysis

We estimated pooled ORs/RRs and standard errors using a random-effects model based on DerSimonian and Laird's approach [14]. This model was adopted because it considers heterogeneity, which was present between our studies. It assumes that, in addition to the presence of random sampling error, any variability in mean effect size is also due to variation in study populations and procedures (between-study heterogeneity).

Due to the low prevalence of RA, which is estimated to affect 0.81% of the UK adult population [15], ORs and RRs were used interchangeably [16]. Individual study and overall effect size data were summarized using forest plots. Case–control and cohort studies were analysed both separately and together, with pooled ORs/RRs calculated in both instances. For all statistical tests $P < 0.05$ were considered significant. Data were analysed using the statistical environment R, version 2.14.1 (R Foundation for Statistical Computing, Vienna, Austria), Stata, version 10.1 (Stata Corp., College Station, TX, USA) and MetaP (Dongliang G, Duke Institute For Genome Sciences & Policy, NC, USA) [17].

Study heterogeneity

Between-study heterogeneity was assessed using Cochran's Q-test and the I^2-statistic. The latter describes the percentage of total variation across studies due to heterogeneity rather than chance. It ranges from 0% (no heterogeneity) to 100% (high heterogeneity) with I^2-values of 25, 50 and 75% having tentatively been suggested to represent low, moderate and high heterogeneity, respectively [18]. Meta-regression was undertaken to evaluate study publication year and quality as potential sources of heterogeneity between studies [19].

Within-study heterogeneity (differences between cases and controls) was evaluated descriptively.

Primary outcome analysis

Our primary outcome measure was the OR of developing RA in drinkers *vs* non-drinkers. Where possible we combined OR/RRs that had adjusted for confounding variables such as age, gender and smoking. In studies only reporting adjusted risks stratified by alcohol intake we combined the different alcohol intake groups into one common OR/RR to estimate the risk of RA for all drinkers using an inverse variance fixed-effects model [19]. In studies not reporting adjusted risks, unadjusted risks were used (calculated from crude data). Due to heterogeneity in the reporting of alcohol intake and RA risk we used the following categories to represent no alcohol in three studies: 0–1 U/week [9], <1 or never glasses of alcohol/week [20] and self-classification as a never-regular drinker [6]. The latter was used because adjusted risks were only reported for never- *vs* ever-regular drinkers and this study had significant differences between cases and controls for age, gender and smoking status, which meant that unadjusted risks for drinkers *vs* non-drinkers calculated from crude data could be affected by confounding.

Subgroup analysis

We undertook three subgroup analyses evaluating: (i) risk differences between ACPA/RF-positive and ACPA/RF-negative RA; (ii) the impact of alcohol quantity on RA risk; and (iii) the impact of drinking duration on RA risk.

Due to variation between studies in the categories of alcohol intake used to report risk it was not possible to combine them to give a summary OR for each drinking category. We therefore examined dose–risk relationships within studies using trend tests. Where these were not reported the Cochran–Armitage test for trend was calculated using crude data [21]. We also combined trend test *P*-values from each study using Stouffer's Z trend test [22]. Levels of alcohol intake were broadly grouped into three categories—low, moderate and high—according to individual study classifications. In studies using low alcohol intake as the reference group, ORs/RRs were recalculated using crude data and taking no alcohol as the reference group. Because only two studies evaluated alcohol intake on more than one occasion (both reporting risk differently) the impact of drinking duration on RA was evaluated descriptively.

Publication bias and selection bias

Publication bias was looked for by constructing funnel plots and applying Begg and Mazumudar's [23] adjusted rank correlation method and Sterne and Egger's [24] linear regression approach.

Examining the influence of individual studies

We repeated our analyses excluding one study at a time. This allowed us to investigate the influence of individual studies on the meta-analysis summary OR and ensure our findings were not attributable to a single study with a large effect size [25].

Results

Studies identified

We screened 893 articles, identifying 22 potential manuscripts from their title or abstract (Fig. 1). Fourteen were excluded: three evaluated the same cohort [2, 26, 27]; two evaluated other risk factors [28, 29]; five examined other issues [30-34]; three were not observational studies (two letters [35, 36] and one review [37]); one did not report effect size data or provide data for its calculation [38]. Additionally the latter study used a self-reported diagnosis of RA that increased the likelihood of case misclassification (reflected in its high reported RA prevalence figures of 5% in men and 7% in women). We therefore did not attempt to contact its authors to obtain raw data for effect size calculation. Eight articles were included, which reported nine separate studies (Table 1). Three were cohort and six were case–control designs.

Cohort studies

The minimum, maximum and median values for the cohort study sizes comprised 18 944; 34 141 and 31 336, respectively. The same values for the ages of included individuals comprised 45, 62 and 61 years, respectively. Two studies examined only females; in the other gender was not described. Two reported serology: one restricted analysis to RF-positive RA; in the other 61% of cases were RF positive. Follow-up periods were 7 [20], 11 [39] and 16 years [40].

Case–control studies

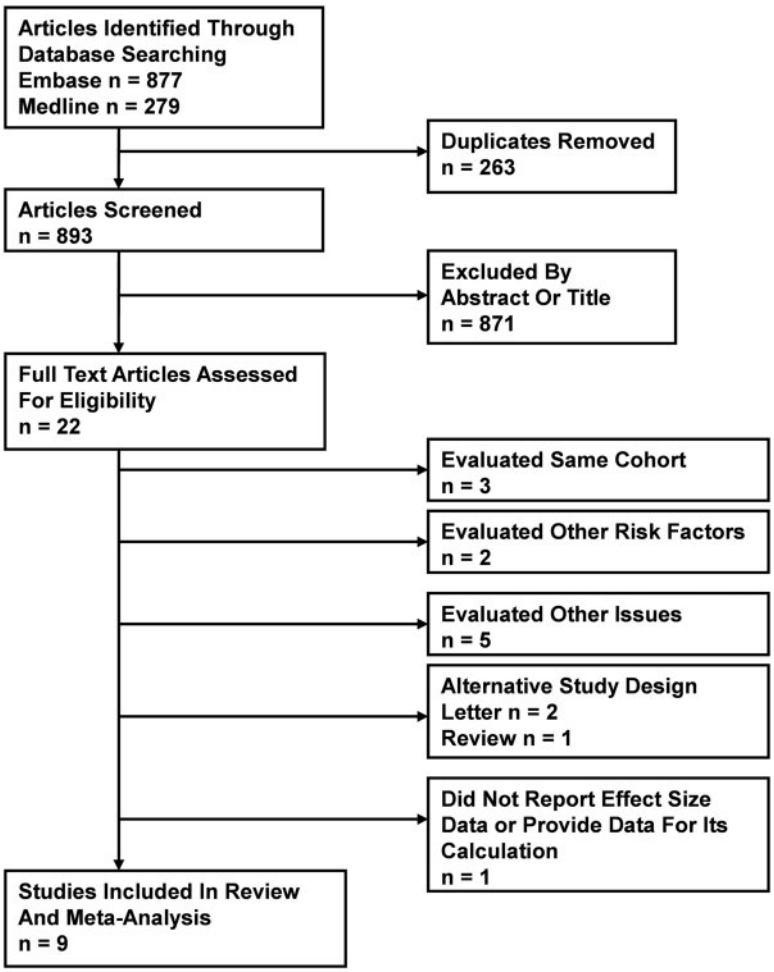
As summary statistics for the case–control studies, we present the minimum, maximum and median number of cases evaluated by each study (135; 1204 and 501.5), the number of controls evaluated (378; 4234 and 937.5), the ages of included cases (49, 61 and 51 years) and the ages of included controls (48, 52 and 50 years). Most studies examined females: three studies evaluated only females; in the remainder 70% were women. One study recruited controls from the same hospital clinic as cases [41]; the remainder recruited population-based controls. Three studies reported disease duration: two evaluated early (⩽2 years) [7] and one established (mean duration 14 years) RA [6]. Four studies reported serology, with seropositive cases ranging from 50 to 79%. Only one reported data on erosions and disease activity: all patients had erosive RA and most moderate disease activity [6].

Alcohol intake definitions and assessment

There was marked variation in the definition of alcohol intake between studies (Table 1), which defined alcohol consumption on a daily, weekly, monthly or lifetime basis and in units, number of drinks or grams of alcohol.

There was further heterogeneity between studies regarding alcohol intake assessment methods. Five used questionnaires [6, 7, 20, 39, 40], three interviews [7, 8, 41] and one medical record review [9]. Alcohol intake was also evaluated at different time points prior to or at RA onset: two cohort studies recorded intake at cohort

FIG. 1 Search strategy.



baseline [39, 40], one study recorded intake 10 years previously [7], two recorded current intake alongside habitual consumption [6, 7], one recorded intake at initial presentation [41], one recorded intake at an unspecified time point [9] and two evaluated intake at multiple time points [8, 20].

Study quality

Study quality scores are shown in Table 1. One study had a Newcastle–Ottawa Score of 6, seven scored 7 and one scored 8. Some cohort studies were not allocated points because they used questionnaires to evaluate drinking; additionally one study failed to adjust for smoking [39] and another had a relatively short follow-up period of 7 years [20]. Some case–control studies were not allocated points because they captured information on alcohol intake from unblinded interviews or questionnaires, did not fully report response rates or had differing response rates between cases and controls.

Although a proportion of studies had differences between cases and controls with regards to age, gender and smoking, they did not lose points for comparability as they adjusted for these factors in their analysis.

Alcohol intake and the overall risk of RA

All studies

All nine studies evaluated alcohol intake and the risk of RA (Table 2; Fig. 2a). Alcohol drinkers were less likely to develop RA, with a significant risk reduction in drinkers *vs* non-drinkers (OR 0.78; 95% CI 0.63, 0.96).

Cohort studies

When restricting our meta-analysis to cohort studies, a non-significant inverse relationship was observed—the OR for RA in drinkers *vs* non-drinkers was 0.91 (95% CI 0.78, 1.07).

Case–control studies

When restricting our meta-analysis to case–control studies a more significant inverse relationship was seen—the OR for RA in drinkers *vs* non-drinkers was 0.70 (95% CI 0.51, 0.95).

Table 1 Studies evaluating the impact of alcohol on the risk of RA

Study	Year	Area	RA diagnosis	Disease status	Size	Alcohol drinker, %	Female, %	Mean age	RA seropositive, %	Alcohol definition	Study quality		
											Selection	Comparability	Exposure/ outcome
Cohort studies													
Cerhan *et al.* [39]	2002	USA	Physician	RA	158	46	100	62	61	Grams of alcohol/day	3	1	3
				No RA	31 178	–							
Di Giuseppe *et al.* [20]	2012	Sweden	National Register	RA	197	89	100	61	–	No. of drinks/week	3	2	2
				No RA	33 944	87							
Heliövaara *et al.* [40]	2000	Finland	Physician	RA	89	55	–	45	100	Grams of alcohol/month	3	2	3
				No RA	18 855	58							
Case–control studies													
Hazes *et al.* [41]	1990	Holland	ARA 1957	Cases	135	23	100	–[a]	–	No. of drinks/day	3	2	2
				Controls	378	36	100	–[a]					
Källberg *et al.*—EIRA [7]	2009	Sweden	ACR 1987	Cases	1204	85	73	51	61	Average no. of drinks/week	4	2	1
				Controls	871	88	74	52					
Källberg *et al.*—CACORA [7]	2009	Denmark	ACR 1987	Cases	444	82	70	49	69	Average no. of drinks/week	3	2	2
				Controls	533	90	61	50					
Maxwell *et al.* [6]	2010	UK	ACR 1987	Cases	873	63	72	61	79	Classification as never-/ever-regular drinker	3	2	1
				Controls	1004	89	65	48					
Rodriguez *et al.* [9]	2009	UK	Physician	Cases	559	43	72	–[a]	–	Units/week	3	2	2
				Controls	4234	43	73	–[a]					
Voigt *et al.* [8]	1994	USA	ARA 1957	Cases	349	84	100	–[a]	50	Lifetime average no. of drinks/week	4	2	1
				Controls	1457	84	100	–[a]					

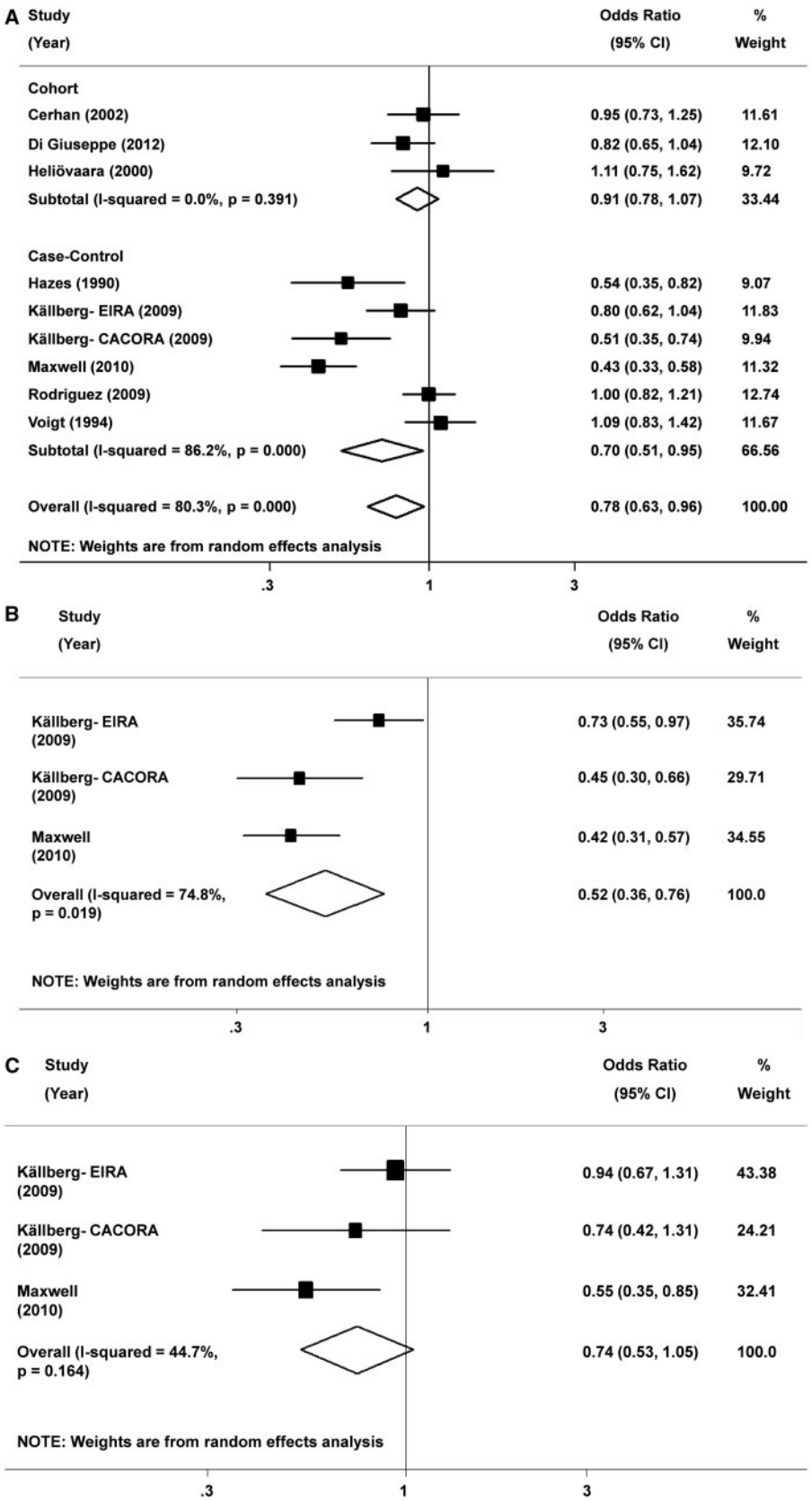
[a]Reported number of cases/controls in different birth year/age groups and not mean age.

Table 2 Alcohol intake and the odds ratios/relative risks (95% CIs) for RA in individual studies

Study	Serology	Alcohol intake				Adjustment factors	Trend test, *P*
		Low	Moderate	High	Any *vs* no alcohol		
Cohort studies							
Cerhan *et al.* [39]	All	0.77 (0.46, 1.27)	1.21 (0.80, 1.83)	0.83 (0.51, 1.36)	0.95 (0.73, 1.25)	Age	0.85
Di Giuseppe *et al.* [20]	All	0.86 (0.56, 1.33)	0.99 (0.67, 1.46)	0.63 (0.42, 0.96)	0.82 (0.65, 1.04)[a]	Age, smoking	0.04
Heliövaara *et al.* [40]	RF+	1.15 (0.64, 2.05)	1.09 (0.59, 2.10)	1.04 (0.44, 2.50)	1.11 (0.75, 1.62)	Age, sex, education, BMI, cholesterol, smoking, coffee intake	0.86[b]
Case–control studies							
Hazes *et al.* [41]	All	–	0.62 (0.40, 0.98)	0.31 (0.13, 0.74)	0.54 (0.35, 0.82)	Birth year, age at RA onset, parity, smoking, OCP use, marital status, menopause	0.10
Källberg *et al.*—EIRA [7]	All	1.02 (0.78, 1.33)	0.59 (0.43, 0.80)	0.56 (0.41, 0.76)	0.80 (0.62, 1.04)	None	<0.0001
	ACPA+	0.94 (0.70, 1.26)	0.50 (0.35, 0.72)	0.52 (0.36, 0.73)	0.73 (0.55, 0.97)		<0.0001
	ACPA−	1.17 (0.83, 1.66)	0.74 (0.49, 1.12)	0.64 (0.42, 0.97)	0.94 (0.67, 1.31)		0.0005
Källberg *et al.*—CACORA [7]	All	0.58 (0.39, 0.87)	0.56 (0.36, 0.86)	0.35 (0.22, 0.55)	0.51 (0.35, 0.74)	None	0.0003
	ACPA+	0.51 (0.33, 0.78)	0.50 (0.32, 0.79)	0.29 (0.18, 0.49)	0.45 (0.30, 0.66)		<0.0001
	ACPA−	0.85 (0.47, 1.56)	0.76 (0.40, 1.46)	0.53 (0.27, 1.06)	0.74 (0.42, 1.31)		0.43
Maxwell *et al.* [6]	All	0.30 (0.23, 0.39)	0.17 (0.12, 0.22)	0.15 (0.11, 0.21)	0.43 (0.33, 0.58)[c]	None	<0.0001[b]
	ACPA+	0.29 (0.22, 0.38)	0.16 (0.11, 0.22)	0.14 (0.10, 0.19)	0.42 (0.31, 0.57)[c]		<0.0001[b]
	ACPA−	0.36 (0.23, 0.56)	0.17 (0.10, 0.30)	0.23 (0.14, 0.38)	0.55 (0.35, 0.85)[c]		<0.0001[b]
	RF+	0.27 (0.21, 0.37)	0.15 (0.10, 0.20)	0.14 (0.10, 0.19)	0.19 (0.15, 0.24)		<0.0001[b]
	RF−	0.33 (0.22, 0.48)	0.23 (0.15, 0.36)	0.22 (0.14, 0.33)	0.26 (0.19, 0.37)		<0.0001[b]
Rodriguez *et al.* [9]	All	0.94 (0.76, 1.17)	1.37 (0.81, 2.33)	1.06 (0.46, 2.45)	1.00 (0.82, 1.21)[d]	Age, sex, year, number of primary care visits/referrals, smoking, BMI, diabetes, CV disease, infections, asthma, anaemia, pregnancy	0.37[b]
Voigt *et al.* [8]	All	1.05 (0.81, 1.36)	0.70 (0.44, 1.13)	0.85 (0.46, 1.57)	1.09 (0.83, 1.42)	Age, smoking, BMI	0.31[b]

[a]<1 glass of alcohol/week used to represent no alcohol. [b]Trend test calculated from crude data. [c]OR for ever- *vs* never-regular drinkers adjusted for age, smoking and gender. [d]0–1 U/week used to represent no alcohol. OCP: oral contraceptive pill; CV: cardiovascular.

Fig. 2 Forest plots of the ORs for RA in alcohol drinkers *vs* non-drinkers.



(**A**) All RA, (**B**) ACPA-positive RA, (**C**) ACPA-negative RA.

Alcohol intake and the risk of ACPA-positive and ACPA-negative RA

Three case–control studies evaluated the impact of alcohol intake on the risk of ACPA-positive and ACPA-negative RA (Table 2; Figs 2b and 2c). There was a significant risk reduction for ACPA-positive RA (OR 0.52; 95% CI 0.36, 0.76). Although there was a lower risk of ACPA-negative RA this was not statistically significant (OR 0.74; 95% CI 0.53, 1.05).

Alcohol intake and the risk of RF-positive and RF-negative RA

Only one case–control study evaluated the impact of alcohol on both RF subsets with significant unadjusted risk reductions for RF-positive (OR 0.19; 95% CI 0.15, 0.24) and RF-negative RA (OR 0.26; 95% CI 0.19, 0.37) [6]. One cohort study evaluated risk in RF-positive patients only with no risk reduction observed (OR 1.11; 95% CI 0.75, 1.62) [40].

Dose–risk relationship between alcohol and RA

This was evaluable in all nine studies (Table 2). Four reported a significant inverse dose–risk relationship between alcohol intake and RA development. The remainder showed no relationship. Combining trend tests across studies provided an estimated summary *P*-value of 0.09; there was therefore no overall significant dose–risk relationship present.

Impact of alcohol intake duration on RA risk

One cohort study evaluated intake at two time points a decade apart, finding that those consuming more than three glasses of alcohol per week sustained over this period were at a lower risk of RA compared with those with less prolonged drinking [20]. One case–control study evaluated average lifetime alcohol intake; no impact on RA risk was observed [8].

Individual study influences

Excluding individual studies from our meta-analysis did not radically alter the summary OR for RA (Table 3). Our findings were therefore not solely attributable to a single study. Excluding three case–control studies resulted in a borderline significant OR (upper 95% CI 1.00).

Study heterogeneity

There was significant heterogeneity between studies examining the overall risk of RA ($P < 0.0001$; $I^2 = 80.3\%$) and ACPA-positive RA ($P = 0.019$; $I^2 = 74.8\%$) but not those evaluating ACPA-negative RA ($P = 0.16$; $I^2 = 44.7\%$).

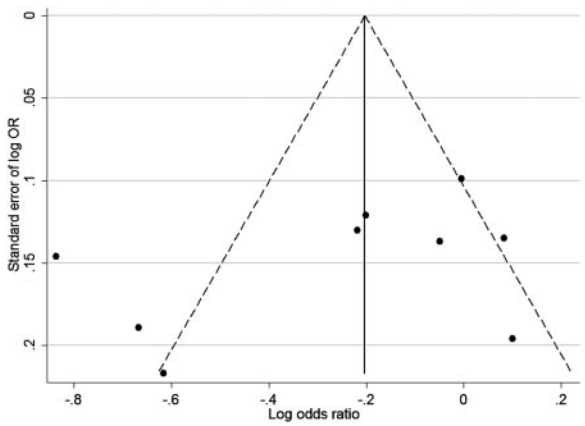
There was no evidence that study publication year ($P = 1.00$) influenced RA risk when evaluated by meta-regression (although the small number of included studies meant this technique had limited power). The impact of study quality was significant ($P = 0.03$), with increasing quality associated with a reduced protective effect of alcohol. This is attributable to the lower score ascribed to Maxwell *et al.* [6] and the higher score ascribed to Heliövaara *et al*. [40], but is difficult to interpret given the score's limited variability.

Three studies reported differences between cases and controls in age and gender: one study had younger controls (average difference 13 years) [6] and two studies had more female cases [6, 7]. Five studies reported differences in smoking status: in four studies smoking rates

Table 3 Individual study influences on RA risk in drinkers *vs* non-drinkers: meta-analysis results with each study omitted

Study	Study design	Summary OR (95% CI) for RA if study omitted
Meta-analysis of all studies		
Cerhan *et al*. [39]	Cohort	0.75 (0.59, 0.96)
Di Giuseppe *et al.* [20]	Cohort	0.77 (0.60, 0.98)
Heliövaara *et al.* [40]	Cohort	0.75 (0.60, 0.94)
Hazes *et al.* [41]	Case–control	0.80 (0.65, 1.00)
Källberg *et al.*—EIRA [7]	Case–control	0.77 (0.60, 0.98)
Källberg *et al.*—CACORA [7]	Case–control	0.81 (0.66, 1.00)
Maxwell *et al.* [6]	Case–control	0.85 (0.71, 1.00)
Rodriguez *et al.* [9]	Case–control	0.75 (0.59, 0.95)
Voigt *et al.* [8]	Case–control	0.74 (0.59, 0.93)
Meta-analysis of studies evaluating ACPA-positive RA		
Källberg *et al.*—EIRA [7]	Case–control	0.43 (0.34, 0.55)
Källberg *et al.*—CACORA [7]	Case–control	0.56 (0.33, 0.96)
Maxwell *et al.* [6]	Case–control	0.58 (0.36, 0.95)
Meta-analysis of studies evaluating ACPA-negative RA		
Källberg *et al.*—EIRA [7]	Case–control	0.61 (0.43, 0.87)
Källberg *et al.*—CACORA [7]	Case–control	0.73 (0.43, 1.24)
Maxwell *et al.* [6]	Case–control	0.88 (0.66, 1.18)

Fig. 3 Funnel plot of the included studies.



were higher in cases [6, 7, 9] and in one study smoking rates were higher in controls [41]. Most studies adjusted for these potentially confounding variables—comprising age, gender and smoking—in multivariate analyses.

Publication bias

There was some evidence of funnel plot asymmetry (Fig. 3), which displays two lower precision studies with large effect sizes favouring a protective effect of alcohol against RA and an absence of similarly lower precision studies of effect sizes in the opposing direction. Although this suggests publication bias may be present, such bias was not detected by Begg's rank correlation ($P = 0.25$) or Egger's weighted regression methods ($P = 0.22$). Additionally the small number of included studies means that publication bias is difficult to assess [25].

Discussion

Our systematic review provides evidence of an inverse relationship between the presence of RA and the consumption of alcohol at or prior to disease onset. It shows that this relationship is predominantly confined to ACPA-positive RA, with a non-significant association observed for ACPA-negative RA. Although these findings suggest that alcohol protects against ACPA-positive RA and support the concept that environmental risk factors differ between RA subsets defined by ACPA status, caution is required in their interpretation as this significant relationship is confined to case–control studies, which have marked heterogeneity between them.

The discrepancy in risk according to ACPA status is interesting. These two subsets are known to differ phenotypically with ACPA-positive RA having lower remission rates and more radiographic erosions [42, 43]. Their underlying genetic and environmental risk factors also appear to differ, which is a concept mirrored by our review. A recent genome-wide association study has indicated distinct genetic architectures with risk allele frequency differences between subsets [44]. A large case–control study has identified divergent environmental risks with ACPA-positive RA linked with smoking, alcohol and oral contraceptive pill use and ACPA-negative RA linked with obesity [2]. There is therefore growing evidence that these subsets have different pathophysiologies and may be considered distinct disease entities [45]. Our review supports this perspective. Due to a lack of data it was not possible to establish if similar environmental risk differences existed for RF-positive and RF-negative RA, although evidence from a smoking meta-analysis suggests these may be present [4].

One potential mechanism through which alcohol could protect against the development of RA is via attenuation of the innate inflammatory response. In experimental animal models, alcohol inhibited the onset of a collagen-induced inflammatory arthritis through down-regulating leucocyte migration, up-regulating testosterone secretion and reducing nuclear factor-κB (NF-κB) activation [46]. Alcohol has also been shown to have anti-inflammatory effects in humans through similar mechanisms such as reducing NF-κB-driven inflammatory mediator production by monocytes [47] (a key cellular pathway in RA [48]). An alternative explanation for this inverse relationship is that individuals with low–moderate alcohol intake have healthier lifestyles compared with complete abstainers who may do so for reasons such as chronic illness; RA could therefore result from confounding lifestyle factors. This explanation has been proposed to explain the J-shaped relationship that exists between drinking alcohol and overall mortality and cardiovascular disease [49].

It would have been of interest to establish if the effects of alcohol on RA risk differ by gender. This is present with smoking, with substantially higher ORs for RA seen in male compared with female smokers [4]. Due to a lack of data we could not establish this for alcohol although the single study that subdivided risk between sexes found similar protective effects in both males and females [6]. Additionally we could not systematically examine the combined effect of smoking and alcohol on RA risk, which was only reported in two studies. These did, however, indicate a likely environment–environment interaction with a greater alcohol-related risk reduction for ACPA-positive RA observed in ever-smokers compared with never-smokers [7]. The importance of adjusting risk for smoking was highlighted by Di Giuseppe *et al*. [20], who found that while drinking was commoner in RA cases because smoking was more prevalent in drinkers the smoking adjusted risk for RA was reduced in those who drank. We consider that as most studies adjusted for smoking status in their analyses, the beneficial effects of alcohol on RA risk were not confounded by smoking.

Our review has several important limitations. Firstly, an overall significant relationship between alcohol and RA was observed in case–control but not cohort studies. Case–control studies are subject to recall bias, which in this context is a distinct possibility. Numerous reasons exist for individuals with RA to consume less alcohol (such as DMARD use), and therefore asking them to

recall past drinking behaviour could be influenced by their current low intake post-diagnosis. Only one case–control study evaluated alcohol intake prior to RA onset; the remainder captured information on alcohol intake at RA development or prior to it through retrospective questionnaires or interviews with no independent means of validating their data. These studies were thus all subject to recall bias. Secondly, due to reporting adjusted risks with low alcohol intake as a reference group, unadjusted risks were used for two studies in the ACPA sub-group analysis [7]. There was, however, little difference between unadjusted and adjusted risks (supplementary Table S3, available as supplementary data at *Rheumatology* Online); we therefore consider that the observed relationship is unlikely to be due to confounding variables. Thirdly, we used self-classification as a never-regular drinker to represent no alcohol intake in one study. Although this is not the same as absolute abstinence the numbers of never-regular drinkers were very similar to those reporting no alcohol intake; this therefore represented an appropriate surrogate measure. Finally, there was significant clinical and methodological heterogeneity between the studies, with important differences existing in how they defined and captured alcohol intake; such heterogeneity limits their suitability to be combined within a meta-analysis.

Further research is needed to establish whether alcohol truly protects against the development of RA. One possible explanation for the lack of an overall association observed in the cohort studies was their failure to evaluate RA cases by ACPA status. Ideally, a large prospective cohort study is required, which subdivides incident cases of RA by the presence or absence of ACPA and captures detailed information on disease risk factors and outcomes. Such an approach could provide crucial insight into many RA risk factors in addition to alcohol. This would enable the development of accurate prediction models combining clinical and genetic risk factors to identify individuals at risk of RA. The end result would be the implementation of evidence-based preventative strategies to halt RA development [50].

Rheumatology key messages

- Case–control studies suggest drinking alcohol may be associated with a reduced risk of RA.
- Alcohol is mainly associated with ACPA-positive RA implying that environmental risks differ by ACPA status.
- Research is required using cohort studies subdividing RA cases by ACPA to determine causality.

Acknowledgements

The authors would like to acknowledge and thank Dr Evangelos Vassos for his comments.

Funding: This work was supported by Arthritis Research UK (19739 to I.C.S.). This manuscript was undertaken as part of an Arthritis Research UK Clinical Research Fellowship (I.C.S.).

Disclosure statement: I.C.S. is undertaking a clinical research fellowship, which is funded by Arthritis Research UK. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

1 Liao KP, Alfredsson L, Karlson EW. Environmental influences on risk for rheumatoid arthritis. Curr Opin Rheumatol 2009;21:279–83.

2 Pedersen M, Jacobsen S, Klarlund M *et al*. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. Arthritis Res Ther 2006;8:R133.

3 Stahl EA, Raychaudhuri S, Remmers EF *et al*. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Gen 2010;42:508–14.

4 Sugiyama D, Nishimura K, Tamaki K *et al*. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 2010;69:70–81.

5 Karlson EW, Chibnik LB, Kraft P *et al*. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann Rheum Dis 2010;69:1077–85.

6 Maxwell JR, Gowers IR, Moore DJ, Wilson AG. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. Rheumatology 2010;49:2140–6.

7 Källberg H, Jacobsen S, Bengtsson C *et al*. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Ann Rheum Dis 2009;68:222–7.

8 Voigt LF, Koepsell TD, Nelson JL, Dugowson CE, Daling JR. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. Epidemiology 1994;5:525–32.

9 Rodriguez LAG, Tolosa LB, Ruigomez A, Johansson S, Wallander MA. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. Scand J Rheumatol 2009;38:173–7.

10 Stroup DF, Berlin JA, Morton SC *et al*. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.

11 Wells GA, Shea B, O'Connell D *et al*. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute Website. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (3 December 2011, date last accessed).

12 Maxwell L, Santesso N, Tugwell PS, Wells GA, Judd M, Buchbinder R. Method guidelines for Cochrane Musculoskeletal Group systematic reviews. J Rheumatol 2006;33:2304–11.

13 Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005;5:19.

14 DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.

15 Symmons D, Turner G, Webb R *et al*. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology 2002;41: 793–800.

16 Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 1998;280:1690–1.

17 Dongliang G. MetaP: a program to combine P values. http://people.genome.duke.edu/~dg48/metap.php (14 July 2012, date last accessed).

18 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557–60.

19 Egger M, Smith GD, Altman DG, eds. Meta-analysis in Stata. Systematic reviews in health care meta-analysis in context. London: BMJ Books 2001;347–69.

20 Di Giuseppe D, Alfredsson L, Bottai M, Askling J, Wolk A. Long term alcohol intake and risk of rheumatoid arthritis in women: a population based cohort study. BMJ 2012;345: e4230.

21 Lachin J. Power and sample size evaluation for the Cochran-Mantel-Haenszel mean score (Wilcoxon rank sum) test and the Cochran-Armitage test for trend. Stat Med 2011;30:3057–66.

22 Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol 2005;18:1368–73.

23 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994; 50:1088–101.

24 Sterne J, Egger M. Publication bias in meta-analysis. prevention, assessment and adjustments. Chichester: John Wiley & Sons, 2005;99–110.

25 Tobias A. sbe26: assessing the influence of a single study in meta-analysis. Stata Tech Bul 1999;47:15–7.

26 Criswell LA, Merlino LA, Cerhan JR *et al*. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. Am J Med 2002;112:465–71.

27 Mikuls TR, Cerhan JR, Criswell LA *et al*. Coffee, tea, and caffeine consumption and risk of rheumatoid arthritis: results from the Iowa Women's Health Study. Arthritis Rheum 2002;46:83–91.

28 Arkema EV, Karlson EW, Costenbader KH. A prospective study of periodontal disease and risk of rheumatoid arthritis. J Rheumatol 2010;37:1800–4.

29 Rosell M, Wesley A-M, Rydin K, Klareskog L, Alfredsson L. EIRA Study Group. Dietary fish and fish oil and the risk of rheumatoid arthritis. Epidemiology 2009; 20:896–901.

30 Bradlow A, Mowat AG. Alcohol consumption in arthritic patients: clinical and laboratory studies. Ann Rheum Dis 1985;44:163–8.

31 Hakoda M, Oiwa H, Kasagi F *et al*. Mortality of rheumatoid arthritis in Japan: a longitudinal cohort study. Ann Rheum Dis 2005;64:1451–5.

32 Jacobsson LT, Knowler WC, Pillemer S *et al*. Rheumatoid arthritis and mortality. A longitudinal study in Pima Indians. Arthritis Rheum 1993;36:1045–53.

33 Myllykangas-Luosujarvi R, Aho K, Kautiainen H, Hakala M. Reduced incidence of alcohol related deaths in subjects with rheumatoid arthritis. Ann Rheum Dis 2000; 59:75–6.

34 Pischon N, Pischon T, Kroger J *et al*. Association among rheumatoid arthritis, oral hygiene, and periodontitis. J Periodontol 2008;79:979–86.

35 Puttonen S, Oksanen T, Vahtera J *et al*. Is shift work a risk factor for rheumatoid arthritis? The Finnish Public Sector study. Ann Rheum Dis 2010;69:779–80.

36 Aho K, Heliovaara M. Alcohol, androgens and arthritis. Ann Rheum Dis 1993;52:897.

37 Sofat N, Keat A. Alcohol intake in rheumatic disease: good or bad? Rheumatology 2002;41:125–8.

38 Paganini-Hill A, Ross RK, Henderson BE. Prevalence of chronic disease and health practices in a retirement community. J Chronic Dis 1986;39:699–707.

39 Cerhan JR, Saag KG, Criswell LA, Merlino LA, Mikuls TR. Blood transfusion, alcohol use, and anthropometric risk factors for rheumatoid arthritis in older women. J Rheumatol 2002;29:246–54.

40 Heliövaara M, Aho K, Knekt P, Impivaara O, Reunanen A, Aromaa A. Coffee consumption, rheumatoid factor, and the risk of rheumatoid arthritis. Ann Rheum Dis 2000;59: 631–5.

41 Hazes JM, Dijkmans BA, Vandenbroucke JP, de Vries RR, Cats A. Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. Ann Rheum Dis 1990;49:980–2.

42 van der Helm-van Mil AHM, Verpoort KN, Breedveld FC, Toes REM, Huizinga TWJ. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. Arthritis Res Ther 2005;7: R949–58.

43 van der Woude D, Syversen SW, van der Voort EIH *et al*. The ACPA isotype profile reflects long-term radiographic progression in rheumatoid arthritis. Ann Rheum Dis 2010; 69:1110–6.

44 Padyukov L, Seielstad M, Ong RTH *et al*. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. Ann Rheum Dis 2011;70:259–65.

45 Daha NA, Toes REM. Rheumatoid arthritis: are ACPA-positive and ACPA-negative RA the same disease? Nat Rev Rheumatol 2011;7:202–3.

46 Jonsson I-M, Verdrengh M, Brisslert M *et al*. Ethanol prevents development of destructive arthritis. Proc Natl Acad Sci USA 2007;104:258–63.

47 Mandrekar P, Catalano D, White B, Szabo G. Moderate alcohol intake in humans attenuates monocyte

inflammatory responses: inhibition of nuclear regulatory factor kappa B and induction of interleukin 10. Alcohol Clin Exp Res 2006;30:135–9.

48 Dichamp I, Bourgeois A, Dirand C, Herbein G, Wendling D. Increased nuclear factor-kappaB activation in peripheral blood monocytes of patients with rheumatoid arthritis is mediated primarily by tumor necrosis factor-alpha. J Rheumatol 2007;34:1976–83.

49 Reynolds K, Lewis B, Nolen JDL, Kinney GL, Sathya B, He J. Alcohol consumption and risk of stroke: a meta-analysis. JAMA 2003;289:579–88.

50 Scott IC, Steer S, Lewis CM, Cope AP. Precipitating and perpetuating factors of rheumatoid arthritis immunopathology: linking the triad of genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis. Best Pract Res Clin Rheumatol 2011;25:447–68.

Supplementary Table 1. Description of Newcastle-Ottawa Scoring System for Case-Control Studies (Wells, et al., 2011)

Scoring Domain	Domain Subscale	Description Of Star/Point Allocation
Selection	Adequacy of case definition	1 star is allocated if the case definition is independently validated (e.g. more than one person is used to extract information or a reference is made to primary record source i.e. medical records). 0 stars are allocated if record linkage is used (e.g. ICD database coding), a self-reported definition is used or no description is given.
	Case representativeness	1 star is allocated if all eligible cases are included with the outcome of interest over a defined time period or in a defined catchment area or in a defined hospital(s) or in a health maintenance organisation or an appropriate sample of those cases (e.g. random sample) is included. 0 stars are allocated if these requirements are not satisfying or not stated.
	Control selection	1 star is allocated if controls are derived from the same community as cases and therefore would have been cases had the outcome of interest been present. 0 stars are allocated if controls are derived from a hospital population or no description is given.
	Control definition	1 star is allocated if it is explicitly stated that controls have no history of the outcome. 0 stars are allocated if no mention of the history of outcome is given.
Comparability	Comparability of cases and controls on the basis of design or analysis	2 stars are allocated for the number of confounding factors that cases and controls are either matched for in the study design or adjusted for in the analysis. 1 star is allocated for controlling for the most important factor. A second star is allocated if the study controls for any additional factors.
Exposure	Ascertainment of exposure	1 star is allocated if a secure record (e.g. surgical records) or structured interview is used where the interviewer is blinded to the case/control status. 0 stars are allocated if an unblinded interview or a written-self report or medical records only are used or no description is given.
	Same method of ascertainment for cases and controls	1 star is allocated if the method of ascertainment is the same for cases and controls. 0 stars are allocated if the method is different.
	Non-response rate	1 star is allocated if the non-response rate is the same for both groups. 0 stars are allocated if the rate is different or not described.

Supplementary Table 2. Description of Newcastle-Ottawa Scoring System for Cohort Studies (Wells et al, 2011)

Scoring Domain	Domain Subscale	Description Of Star/Point Allocation
Selection	Exposed cohort representativeness	1 star is allocated if the exposed cohort is representative of the average individual at risk of RA in the community. 0 stars are allocated if the exposed cohort is a selected group of individual's e.g. nurses or no description of the derivation of the cohort is given.
	Non-exposed cohort selection	1 star is allocated if the non-exposed cohort is drawn from the same community as the exposed cohort. 0 stars are allocated if it is drawn from a different source or no description of its derivation is provided.
	Ascertainment of exposure	1 star is allocated if a secure record (e.g. surgical records) or structured interview is used. 0 stars are allocated if a written self-report is used or no description is given.
	Demonstration that outcome of interest was absent at the start of study	1 star is allocated if this is demonstrated. 0 stars are allocated if this is not demonstrated.
Comparability	Comparability of cohorts on the basis of the design or analysis	2 stars are allocated for the number of confounders that the exposed and non-exposed individuals are matched for and/or adjusted for in the analysis. 1 star is allocated for controlling for the most important factor. A second star is allocated if the study controls for any additional factors.
Exposure	Outcome assessment	1 star is allocated if a blind independent assessment is undertaken or record linkage is used. 0 stars are allocated if the assessment is self-reported or no description is given.
	Was follow-up long enough for outcomes to occur	1 star is allocated if an adequate follow-up length is used (in the case of our systematic review we considered 10 or more years to be an adequate follow-up period). 0 stars are allocated if the follow-up length is inadequate.
	Adequacy of follow-up cohorts	1 star is allocated if follow-up is complete and all subjects are accounted for or if a small number of subjects are lost to follow-up, which is unlikely to introduce bias (in the case of our systematic review we considered ≥90% follow-up to be adequate), or a description is provided of those lost. 0 stars are allocated if the follow-up rate is less than this amount and no description is given of those lost or no statement regarding follow-up is given.

Supplementary Table 3. Comparison of Adjusted and Unadjusted Odds Ratios for RA in Drinkers vs. Non-Drinkers from the EIRA and CACORA Studies

Alcohol Intake	EIRA		CACORA	
	Adjusted OR (95% CI)	**Unadjusted OR (95% CI)**	**Adjusted OR (95% CI)**	**Unadjusted OR (95% CI)**
		RA Overall		
Non-Drinkers	1.1 (0.8-1.4)	1.0 (0.8-1.3)	1.7 (1.1-2.5)	1.7 (1.2-2.5)
Low	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)
Mod	0.6 (0.4-0.7)	0.6 (0.4-0.7)	1.0 (0.7-1.3)	1.0 (0.7-1.3)
High	0.5 (0.4-0.6)	0.5 (0.4-0.7)	0.6 (0.4-0.9)	0.6 (0.4-0.8)
		ACPA-positive RA		
Non-Drinkers	1.3 (1.0-1.8)	1.1 (0.8-1.4)	2.0 (1.3-3.1)	2.0 (1.3-3.0)
Low	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)
Mod	0.5 (0.4-0.7)	0.5 (0.4-0.7)	1.0 (0.7-1.4)	1.0 (0.7-1.4)
High	0.5 (0.3-0.6)	0.5 (0.4-0.7)	0.6 (0.4-0.9)	0.6 (0.4-0.9)
		ACPA-Negative RA		
Non-Drinkers	0.9 (0.6-1.3)	0.9 (0.6-1.2)	1.1 (0.6-2.1)	1.2 (0.6-2.2)
Low	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)	1.0 (Reference)
Mod	0.7 (0.5-0.9)	0.6 (0.5-0.9)	0.9 (0.6-1.5)	0.9 (0.6-1.4)
High	0.5 (0.4-0.7)	0.5 (0.4-0.7)	0.9 (0.5-1.5)	0.6 (0.4-1.1)

EIRA = Epidemiological Investigation of Rheumatoid Arthritis; CACORA = CAse-COntrol study on Rheumatoid Arthritis (Kallberg et al, 2009).

CHAPTER 3. GENETIC SUSCEPTIBILITY AND X-RAY DAMAGE

3.1. BACKGROUND

Rheumatoid arthritis (RA) is a heterogeneous disease whose course varies substantially between patients. Biomarkers that reliably identify patients with a severe, progressive phenotype at disease onset are therefore highly desirable; they could be used to inform decisions on treatment intensity. The moderate heritability of radiological progression in RA, which is estimated at 45 to 58% (Knevel et al, 2012b), suggests genetic markers could be important prognostic predictors. Despite success in delineating the genetic architecture of RA susceptibility (Okada et al, 2013, Raychaudhuri et al, 2012) only a few replicated loci have been linked to radiological progression. These have mainly been identified through candidate gene studies with a lack of large, genotyped, longitudinal datasets limiting the power of individual cohorts to identify loci at genome-wide significance.

Existing reports of genetic predictors of X-ray damage in RA have assessed observational study cohorts (Knevel et al, 2013a, Krabben, et al., 2013). These comprise patients with a range of disease activities managed using a variety of treatment strategies. As both of these factors have major impacts on radiological progression, these analyses are unable to differentiate between genetic variants mediating X-ray damage indirectly via influences on disease activity or treatment effects and those having a direct effect. Additionally, as the evidence base for most RA treatments is from RCTs of patients with severe, active disease for genetic predictors to be able to inform clinical decisions they too require evaluation in such individuals.

Several studies have examined the hypothesis that RA susceptibility variants associate with radiological progression. Their findings are often not replicated (Marinou, et al., 2010, Knevel et al, 2012a). Additionally they have examined only a proportion of risk loci (the largest analysis assessed 37 independent variants (De Rooy et al, 2013b)) and evaluated each locus separately, despite limited power.

Recent meta-analyses have expanded the number of non-MHC loci, *HLA-DRB1* alleles and HLA amino acid polymorphisms associated with RA susceptibility (Okada et al, 2013, Raychaudhuri et al, 2012). The aim of this study was to establish

if these associated with radiological progression in early, active RA patients enrolled to two RCTs − the Combination Anti-Rheumatic Drugs in Early RA (CARDERA) trials − which constitute the CARDERA Genetics Cohort (CGC). Their association was evaluated both individually and cumulatively, using a weighted genetic risk score (wGRS) combining risk loci. This study represents the first analysis of genetic associations with radiological progression in early, active RA patients within a clinical trial setting**.**

3.2. METHODS

3.2.1. Ethical Approval

The CARDERA-1 trial was approved by the South Thames Multicentre Research Ethics Committee (REC reference: MREC (1) 99/04). The CARDERA-2 trial was approved by the South East Research Ethics Committee (REC reference: MREC 02/1/089). Approval was obtained to genotype archived DNA from the East of England - Essex Research Ethics Committee (REC reference: 11/EE/0544). All patients provided consent according to the Declaration of Helsinki.

3.2.2. Subjects

3.2.2.1.CARDERA-1

CARDERA-1 recruited 467 patients with early, active RA from multiple UK centres (Choy et al, 2008). Patients were randomised to one of four treatment arms: (1) monotherapy with methotrexate; (2) double therapy with methotrexate and ciclosporin; (3) double therapy with methotrexate and prednisolone; (4) triple therapy with methotrexate, ciclosporin and prednisolone. A factorial-design was adopted to allow the simultaneous evaluation of prednisolone and ciclosporin in a 2 x 2 design. Patients were treated and followed-up for 2 years.

3.2.2.2.CARDERA-2

CARDERA-2 recruited 167 patients with early, active RA from multiple UK centres (Scott and Choy, 2007). Patients were randomised to one of two treatment arms: (1) monotherapy with methotrexate; (2) double therapy with anakinra (an interleukin-1 receptor antagonist) and methotrexate. The randomised treatments were given for 12

months; thereafter patients were treated at the discretion of their hospital rheumatologist. Patients were followed-up for 2 years.

3.2.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for both trials were identical. Inclusion criteria comprised active RA (defined by having 3 out of ≥3 swollen joints, ≥6 tender joints, ≥45 minutes morning stiffness, ESR ≥28mm/hr), which fulfilled the 1987 ACR classification criteria. Patients had to be able to give informed consent and be aged ≥18 years. Exclusion criteria comprised other inflammatory arthropathies; previous methotrexate treatment; contraindications/intolerance of any of the trial drugs; other serious medical disorders; acute/chronic infection; significant neutropenia ($<1.5x10^{12}$/dl); significant thrombocytopenia ($<100x10^{12}$/dl); abnormal liver tests (GGT or AST/ALT >3 times or >2 times the upper limit of normal, respectively); severe renal impairment (creatinine clearance <30ml/min); and the use of low-dose oral steroids for the treatment of RA.

3.2.4. Radiological Assessments

In CARDERA-1 hand and feet radiographs were read for modified Larsen scores at baseline and 6-monthly for 2 years. They were assessed independently by two experienced observers after preliminary studies ensured comparable scoring (Choy et al, 2008). In CARDERA-2 hand and feet radiographs were read for modified Larsen scores at baseline and 12-monthly for 2 years. They were assessed by an experienced observer with specific expertise in RA radiological scoring; this was one of the observers that read X-rays in CARDERA-1.

3.2.5. Genotyping

DNA was genotyped on the Illumina ImmunoChip. The probes on this microarray (designed in 2009) interrogate 195,806 SNPs and 718 small insertion–deletions (Parkes et al, 2013). This chip provides dense coverage of GWAS confirmed immune-mediated disease susceptibility loci. Although it does not provide true genome-wide coverage, it is ideal for examining known immune disease related

regions. Genotyping was undertaken by staff in the Biomedical Research Centre Genomics laboratory (King's College London).

As the ImmunoChip contains a large number of low-frequency, rare variants genotype calling was performed using optiCall (Shah, et al., 2012). This algorithm uses SNP-wise and sample-wise calling to more accurately ascertain genotypes at rare, low-frequency and common variants. Genotype calling was undertaken by Dr Sarah L Spain.

3.2.6. Quality Control Procedures

Individual and marker-level QC procedures are described in Table 3-1. Individuals were removed with <95% SNP call rates (8 patients), anomalously high or low heterozygosity rates defined as a Wright's inbreeding coefficient F of less than or greater than 3 standard deviations from the mean, respectively (2 patients; Figure 3-1), non-European ancestry (20 patients; identified through principal components analysis (PCA) using imported HapMap 3 data (undertaken by Dr Sarah L Spain); Figure 3-1), discordance between genotype and phenotype gender data (2 patients) and duplicate samples (4 patients). Markers were removed that had >5% missingness (3,470 SNPs), were not present in the Genome Reference Consortium Human Build 37 or were duplicate markers (1,748 SNPs), were not in Hardy-Weinberg Equilibrium (HWE) (P<0.00001; 3,158 SNPs) or had low MAFs of <0.01 (49,275 SNPs).

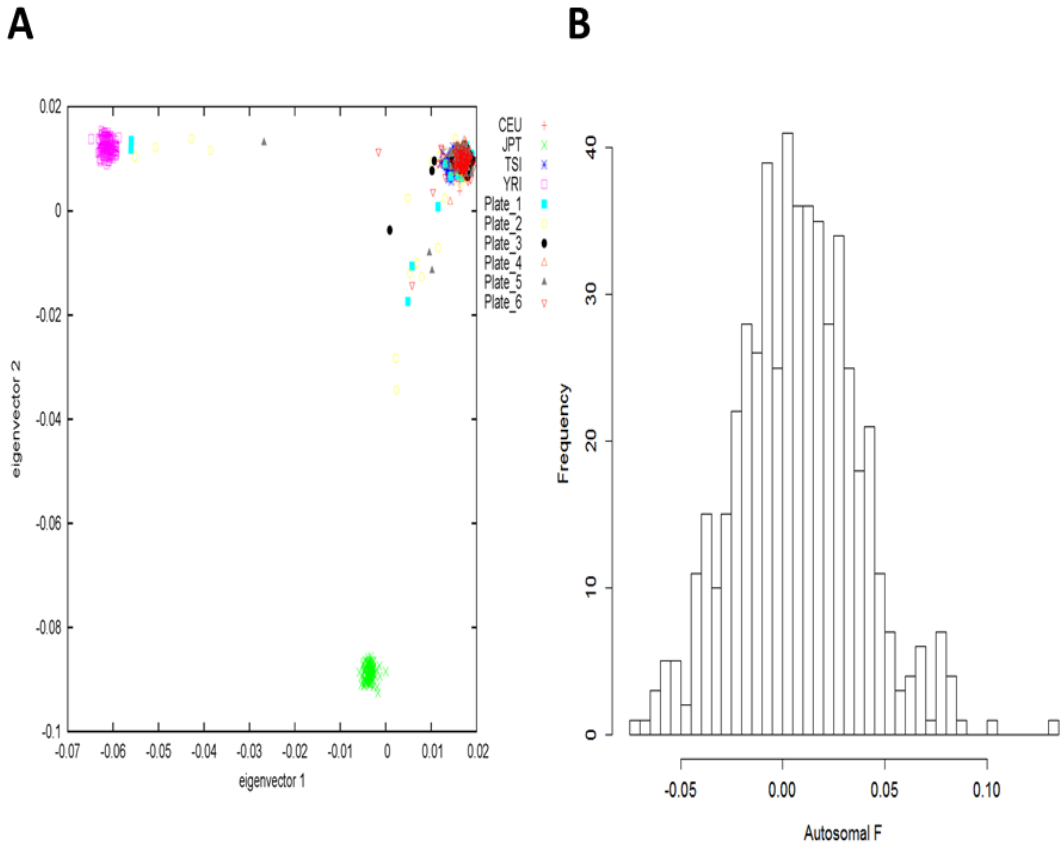
From 560 genotyped individuals, 524 were included in the final analysis (424 from CARDERA-1; 100 from CARDERA-2). From 196,524 genotype markers, 138,873 passed QC procedures (genotyping rate 0.999).

Table 3-1. Genotyping Quality Control Procedures

Individual-Level QC		
QC Stage	**No. Individuals In Dataset**	**No. Individuals Removed**
Pre-QC	560	-
<95% SNP call rates	552	8
Anomalously high/low heterozygosity	550	2
Non-European ancestry[a]	530	20
Relatedness	530	0
Genotype-phenotype gender discordance	528	2
Duplicate samples	524	4
Post-QC	524	-
Marker-Level QC		
QC Stage	**No. SNPs In Dataset**	**No. SNPs Removed**
Pre-QC	196,524	-
Marker missingness >5%	193,054	3,470
Not in GRCh37 assembly/duplicate markers	191,306	1,748
HWE (P<0.00001)	188,148	3,158
MAF < 0.01	138,873	49,275
Post-QC	138,873	-

a= Ethnic outliers identified by PCA plots of samples merged with HapMap3 data; QC = quality control; No. = number; GRCh37 = Genome Reference Consortium Human Build 37; MAF = minor allele frequency; HWE = Hardy-Weinberg equilibrium.

Figure 3-1. Population Outlier and Heterozygosity Assessments



Panel A = Identification of population outliers using principal components analysis (PCA); samples were clustered by ancestry based on the first and second eigenvectors; HapMap 3 data files provided high density coverage of CEU (Utah residents with Northern and Western European ancestry from the CEPH collection), JPT (Japanese in Tokyo, Japan), TSI (Tuscans in Italy) and YRI (Yoruba in Ibandan, Nigeria) populations. Plates 1 to 6 refer to the ImmunoChip plates used to genotype CARDERA-1 and CARDERA-2 patients. Analysis performed by Dr Sarah L Spain.

Panel B = Histogram of the inbreeding coefficient F; 2 patients were removed from the analysis due to F levels >3 standard deviations from the mean indicating excessive homozygosity.

3.2.7. HLA Imputation

Classical *HLA-DRB1* alleles and amino acid polymorphisms in HLA-DRβ1, HLA-B and HLA-DPβ1 molecules were imputed using SNP2HLA (Jia, et al., 2013), which uses the Beagle method for imputation. This was undertaken by Mr Jelmar Quist. Reference data from the Type 1 Diabetes Genetics Consortium (T1DGC) panel was used (Brown, et al., 2009); this contains 5,868 SNPs (genotyped using the ImmunoChip) that extensively tag the MHC region alongside classical four-digit

resolution *HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1* and *-DRB1* alleles in 5,225 unrelated European ancestry individuals (Jia et al, 2013). In the absence of HLA tissue typing data in the CGC it was not possible to validate imputation accuracy within this cohort. However, SNP2HLA has been extensively validated in several large datasets (Han et al, 2014, Jia et al, 2013). Its precision was demonstrated in 918 individuals from the 1958 birth cohort, in whom it was 96.7% and 99.3% accurate at imputing four-digit resolution *HLA-DRB1* alleles and amino acid polymorphisms within HLA genes using ImmunoChip genotypes, respectively (Jia et al, 2013).

3.2.8. Susceptibility Variants Evaluated

Confirmed European ancestry RA susceptibility SNPs from the most recent meta-analysis by Okada *et al* (Okada et al, 2013) were evaluated for their association with radiological progression. Of 77 potential SNPs, 69 were available in the CGC (including 31 proxy SNPs with an $r^2>0.8$; Table 3-2). For the *HLA-DRB1* locus, rs660895 from the Eyre *et al* meta-analysis (Eyre et al, 2012) was used, as the most strongly associated SNP identified by Okada *et al* (rs9268839) was unavailable in the CGC. Nine SNPs had missing data. These comprised rs11702844, rs11933540, rs1893592, rs2306627, rs2834512, rs45475795, rs9979383 (all missing 1 genotype), rs12108065 (4 genotypes missing) and rs9427372 (21 genotypes missing). These missing genotypes were imputed by assigning them the expected allele count, which is twice the allele frequency, for each SNP. This procedure is commonly used to account for missing genotype data (Karlson et al, 2010, Yarwood et al, 2013).

Seventeen four-digit *HLA-DRB1* alleles attaining $P_{GWAS}<5x10^{-8}$ in European ancestry individuals from the most recent RA meta-analysis of association in the HLA region (Raychaudhuri et al, 2012) were evaluated. Amino acids at positions 11, 13, 71 and 74 in HLA-DRβ1, position 9 in HLA-B and position 9 in HLA-DPβ1 were also assessed. In the HLA meta-analysis these explained most of the region's association with RA (Raychaudhuri et al, 2012). Positions 11 and 13 in HLA-DRβ1 were both included as their strong linkage disequilibrium (LD) meant causality could not be assigned (Raychaudhuri et al, 2012).

Table 3-2.Proxy SNPs Used in Analysis

Meta-Analysis SNP				Proxy SNP			
SNP	**Chr**	**BP Position**	**A1/ A2**	**SNP**	**BP Position**	**A1/ A2**	**r^2**
chr1:2523811	1	2523811	G/A	rs10752747	2524915	G/T	0.875
chr17:38031857	17	38031857	G/T	rs10852935	38031674	T/C	0.979
rs10790268	11	118729391	G/A	rs10892299	120000000	C/T	0.915
rs71508903	10	63779871	T/C	rs11593907	63786554	C/T	0.865
rs73194058	21	34764288	C/A	rs11702844	34759876	A/G	0.970
rs4452313	3	17047032	T/A	rs12108065	17074871	G/A	0.925
rs2105325	1	173349725	C/A	rs1557121	173353881	C/T	0.986
rs6715284	2	202154397	G/C	rs16837131	202173812	A/C	0.972
rs2736337	8	11341880	C/T	rs2061831	11339882	C/T	0.993
rs5987194	X	153301467	C/G	rs2075596	153297392	A/G	0.989
rs331463	11	36501787	T/A	rs2303439	36514290	C/T	0.944
rs28411352	1	38278579	T/C	rs2306627	38260503	T/C	0.844
chr21:35928240	21	35928240	C/T	rs2834512	35911599	G/A	0.979
rs8083786	18	12881361	G/A	rs34846641	12875975	G/A	1.000
chr7:128580042	7	128580042	G/A	rs3778754	128575552	G/C	0.948
rs3824660	10	8104722	C/T	rs3802604	8102272	G/A	0.926
rs508970	11	60906450	A/G	rs595158	60909581	A/C	0.929
rs10774624	12	111833788	G/A	rs653178	112007756	T/C	0.861
rs1516971	8	129542100	T/C	rs6651252	129567181	T/C	0.985
rs1633360	12	58108052	T/C	rs701006	58106836	G/A	0.979
rs773125	12	56394954	A/G	rs705700	56389293	T/C	0.967
rs706778	10	6098949	T/C	rs7073236	6106552	C/T	0.910
rs10175798	2	30449594	A/G	rs7579944	30445026	C/T	0.940
rs9603616	13	40368069	C/T	rs7993214	40350912	C/T	0.936
rs2234067	6	36355654	C/A	rs879036	36349890	C/T	0.964
rs12140275	1	38633879	A/T	rs883220	38616871	C/A	0.962
rs1950897	14	68760141	T/C	rs911263	68753593	T/C	0.994
rs9372120	6	106667535	G/T	rs9386514	106636902	C/T	0.992
rs72717009	1	161405053	T/C	rs9427372	161399920	C/T	0.966
rs9826828	3	136402060	A/G	rs9858105	136219264	C/T	1.000
rs8133843	21	36738242	A/G	rs9979383	36715761	T/C	0.894

Proxy SNPs obtained using 1,000 Genomes European population panels (CEU/TSI/GBR/FIN/IBS); BP position = base pair position based on GRCh37 assembly; Chr = chromosome; reference meta-analysis by Okada et al (Okada et al, 2013).

3.2.9. Individual Variant Associations with X-ray Progression

3.2.9.1.Linear Mixed-Effects Model Using Repeated Measures

The association between Larsen score progression and each susceptibility variant (SNP, *HLA-DRB1* allele and HLA amino acid) was examined using a linear mixed-effects model. This assessed all repeated Larsen scores simultaneously and, through incorporating correlated random-effects, accounted for within-individual correlations in Larsen scores over time. This widely used approach for modelling repeated measurements is preferable to evaluating genetic associations with the 24-month change in Larsen scores as a single variable; by using repeated scores it provides a more definitive evaluation of within-individual changes over time and optimises the power to detect significant predictor variables (Guo, et al., 2013). Furthermore, this model could accommodate missing Larsen score data, by making the assumption that the reason they were missing was unrelated to the actual missing data values, termed "missing-at-random" (MAR) (Higgins and Green, 2011). This assumption was reasonable with most missing data occurring in CARDERA-2 patients at 6 and 18 months as it was not study protocol for X-rays to be performed at these time points.

As Larsen scores had a positively-skewed distribution they were log-transformed with a constant added (due to zero unit scores) in order to approximate a normal distribution (Figure 3-2). Log_e (Larsen score + 1) was used as the response variable. Time (years), clinical variables, genotype and a genotype*time interaction term, were included as fixed-effects predictor variables. The clinical variables comprised treatment (categorically coded in 5 levels with methotrexate monotherapy as the reference group), age, disease duration and RF status. These were chosen using stepwise backwards elimination in a model containing only clinical factors as predictor variables; clinical factors resulting in a model with a significantly lower Akaike Information Criterion (AIC) were included. RF status was used instead of ACPA status as ACPA data were missing in 25 patients. These two variables had modest correlation (57% of patients were ACPA-positive/RF-positive; 13% were ACPA-positive/RF-negative; 11% were ACPA-negative/RF-positive; 19% were ACPA-negative/RF-negative; phi coefficient=0.43). Any significant findings were reanalysed using a model containing ACPA in place of RF and omitting the 25 individuals with unknown ACPA status. Significant genetic markers were also

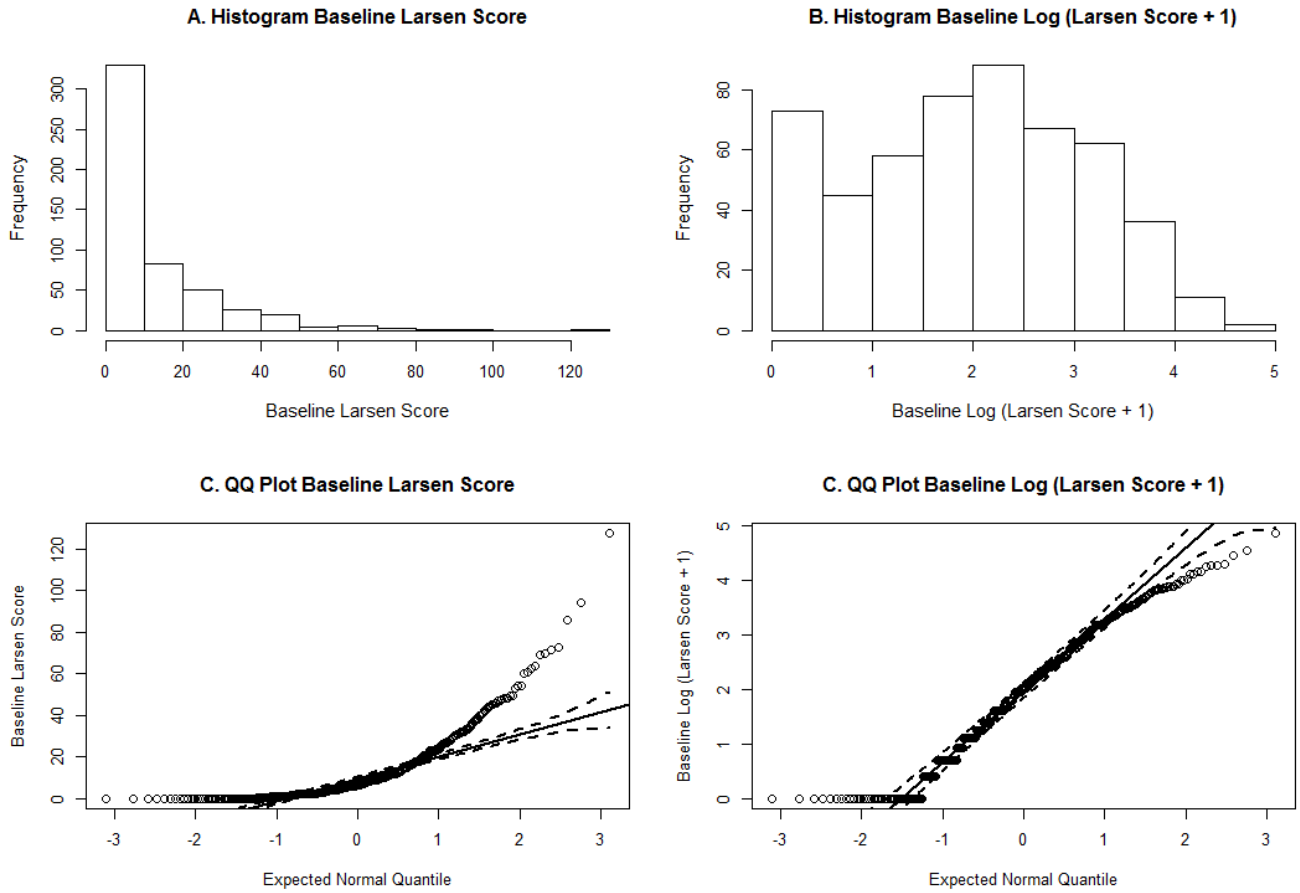
evaluated in ACPA-positive and ACPA-negative patients separately, and the resultant β-values compared to those obtained in all RA patients. This established if their effects were independent of ACPA. β-values, as opposed to *P*-values, were considered in the ACPA subgroup analysis as the power was substantially lower due to the smaller sample size of each group (350 and 149 patients were ACPA-positive and ACPA-negative, respectively) and low X-ray progression rates in ACPA-negative patients (24-month mean Larsen score changes 3.22 (95% CI 1.96-4.48) in ACPA-negative RA vs. 5.88 (95% CI 4.94-6.81) in ACPA-positive RA).

Examination of residuals from a model containing time and clinical variables only confirmed a good fit of the model. Residual plots demonstrated a random distribution of points around the zero line; residuals were normally distributed (Figure 3-3). Using log-transformed Larsen scores therefore fulfilled the assumptions of the model. Genotype was coded additively. All β-estimates were back-transformed to the original Larsen scale. The genotype*time term β provided information on the annual increase in Larsen score per risk allele copy relative to the reference genotype; *P*-values for this term provided information on the variant's role in radiological progression. The X-chromosome SNP (rs5987194) was assessed separately in males and females.

3.2.9.2.Three-Way ANOVA Model Evaluating Change in Larsen Scores

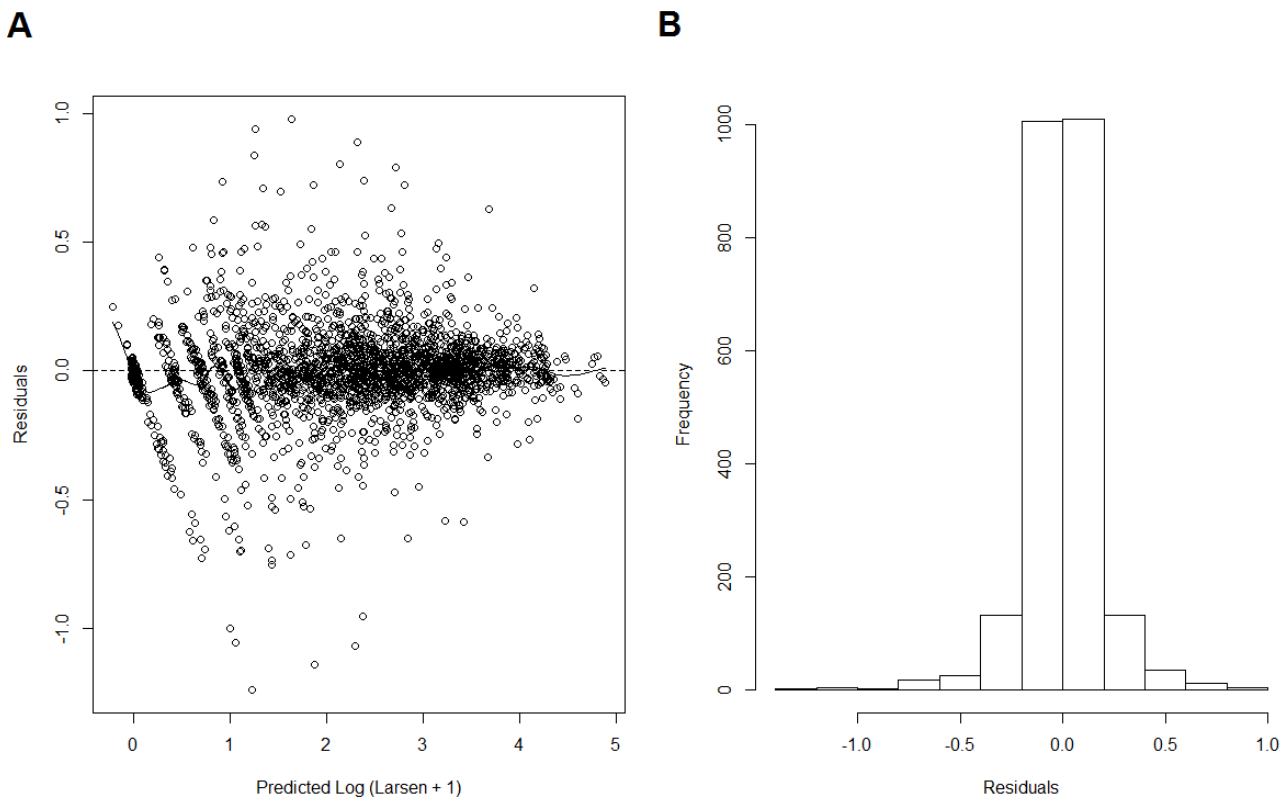
To ensure that the linear mixed-effects model results were not biased by missing Larsen scores (missing at 0, 6, 12, 18 and 24 months in 4, 103, 13, 103 and 19 patients, respectively) an alternative modelling approach was also undertaken. This involved a 3-way ANOVA model including the 24-month change in Larsen scores (ΔLarsen) as the response variable and treatment, RF-status and genotype (coded additively) as predictor variables. As with the linear mixed-effects model relevant clinical variables were chosen through stepwise backwards elimination. As ΔLarsen remained non-normally distributed despite log-transformation, a rank-based inverse normal transformation (INT) of this variable was undertaken using the Blom method (Beasley, et al., 2009) (Figure 3-4). Residual plots confirmed the assumptions of the model were fulfilled (Figure 3-5). The 19 individuals without 24-month X-ray data were excluded from this analysis.

Figure 3-2. Log-Transformation of Baseline Larsen Scores



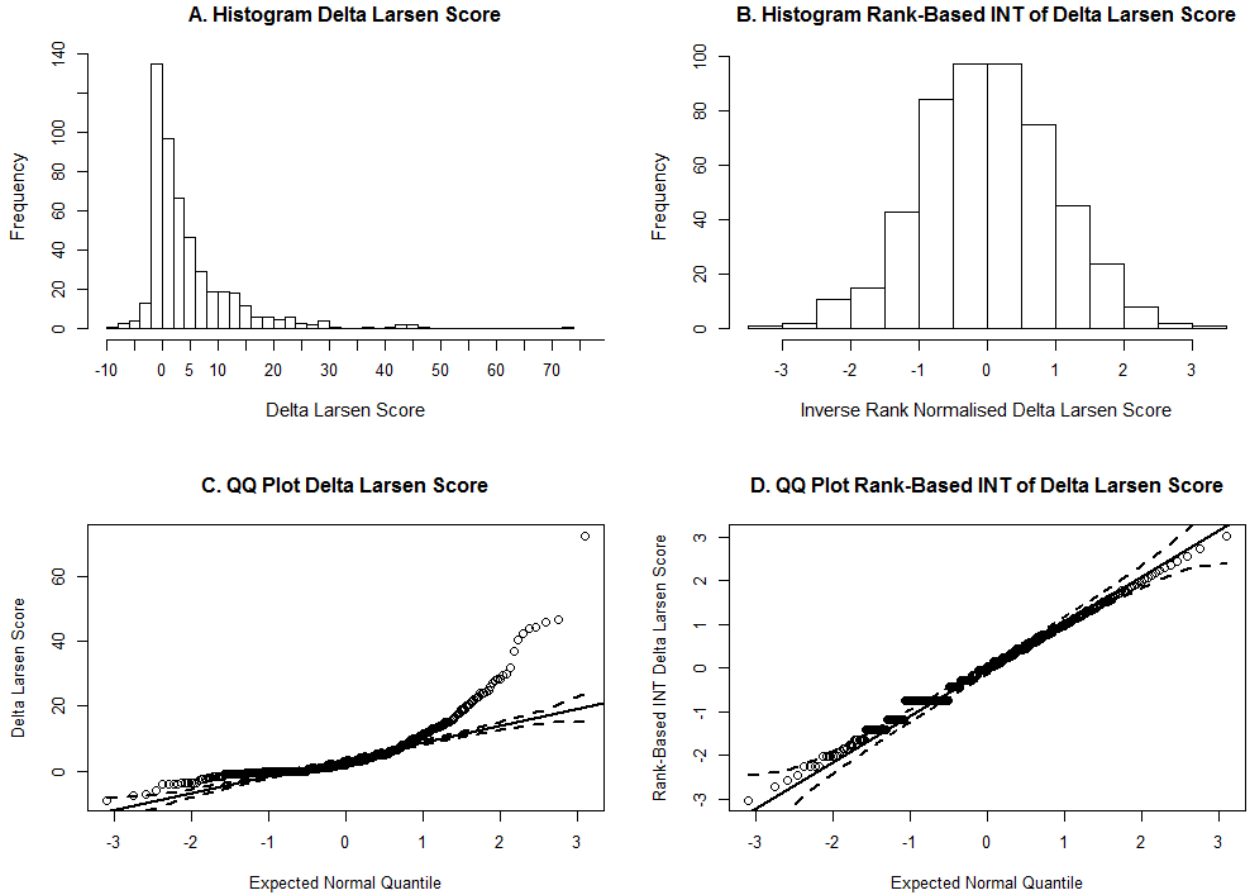
QQ = quantile-quantile; dotted lines in QQ plots represent 95% confidence intervals; similar findings seen for Larsen score transformations at other time points.

Figure 3-3. Analysis of Residuals from a Linear Mixed-Effects Model Including Time and Clinical Variables Only



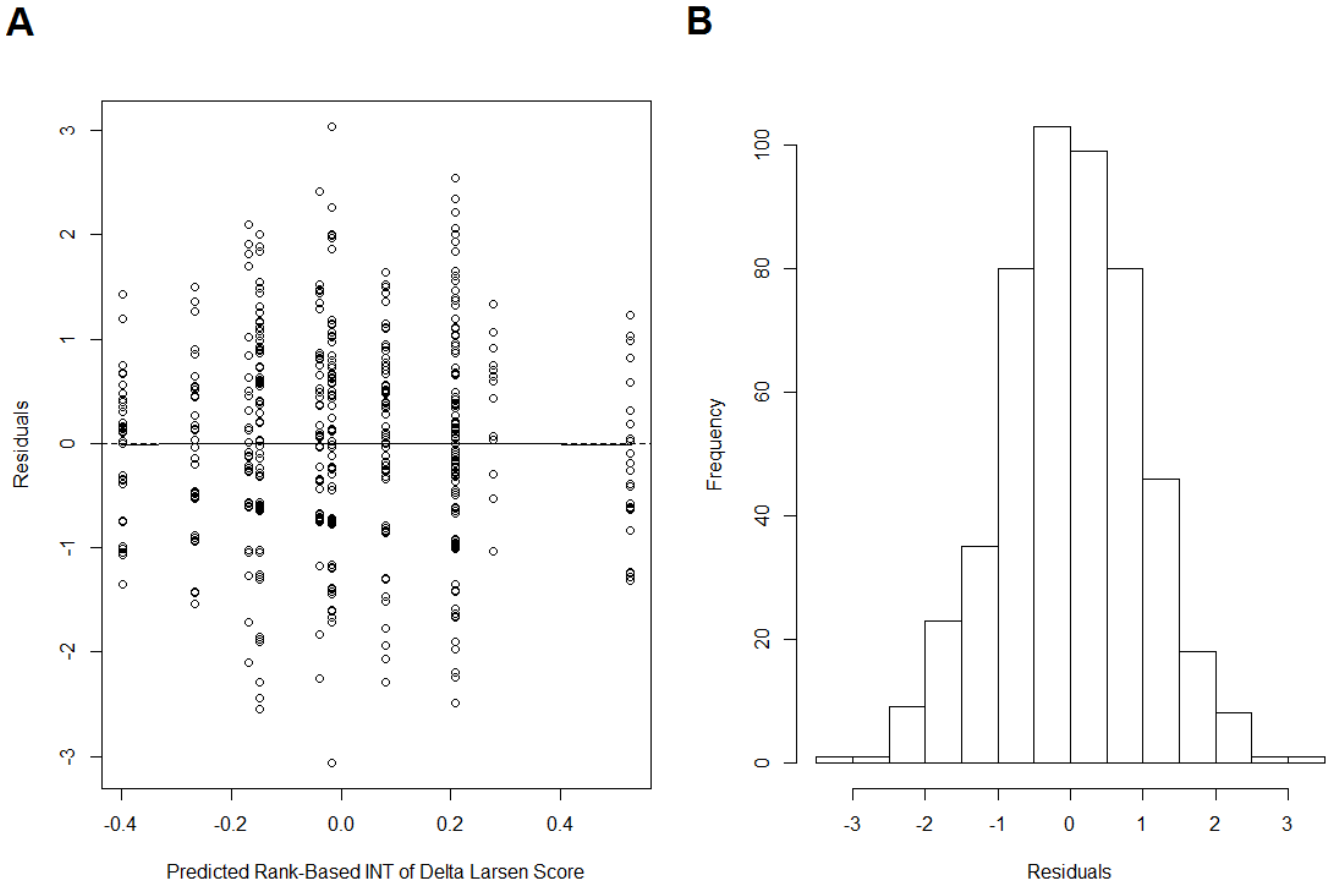
A = Plot of the residuals (difference between observed and predicted $\log_e$(Larsen + 1)) against the predicted $\log_e$(Larsen + 1); dotted line represents the zero point on the y-axis (the smoothing line approximately overlies the dotted line).
B = Histogram of residuals.

Figure 3-4. Rank-Based Inverse Normal Transformation of the 24-Month Change in Larsen Scores



Delta Larsen Score = 24-month change in Larsen score; INT = inverse normal transformation; QQ = quantile-quantile; dotted lines in QQ plots represent 95% confidence intervals.

Figure 3-5. Analysis of Residuals from an ANOVA Model Including Clinical Variables Only



A = Plot of the residuals (difference between observed and predicted inverse rank normalised Larsen score) against the predicted score; dotted line represents the zero point on the y-axis (the smoothing line approximately overlies the dotted line); INT= inverse normal transformation. B = Histogram of residuals.

3.2.10. Weighted Genetic Risk Score Association with X-Ray Progression

The association between a wGRS, derived from all susceptibility loci, and Larsen score progression was tested. Two different wGRSs were evaluated. The first (full wGRS) included all SNPs (excluding the X-chromosome marker); the second (non-HLA wGRS) also excluded the *04:01 tagging SNP.

Both wGRSs were constructed using the R package, REGENT (Risk Estimation for Genetic and Environmental Traits) (Scott et al, 2013b, Crouch, et al., 2013). This uses data on allelic ORs from reference meta-analyses to generate a genotype relative risk score (GRR) across loci within a multiplicative model. In this analysis, the Log_e (GRR) represented the wGRS. Log-transforming the GRR (which is equivalent to an individual's overall OR for RA) ensured the wGRS distribution was symmetrical around 1.00; a non-transformed GRR is limited for scores <1.00, as it cannot have a negative value, resulting in a skewed distribution (Bland and Altman, 2000).

The wGRS replaced the genotype term in the linear mixed-effects and ANOVA models. Data for the genetic loci were obtained from the combined stage of the reference meta-analysis (Okada et al, 2013).

3.2.11. Principal Components Analysis

To account for population stratification, PCA was undertaken using EIGENSOFT (performed by Professor Cathryn Lewis). Analysis of the first 10 principal components (PCs) did not demonstrate any identifiable structure in the CGC. This was confirmed by running a preliminary GWAS evaluating all 138,873 genetic markers as predictors of the rank-based INT ΔLarsen in a simple regression model including genotype (coded additively) alongside the first 3 PCs as predictor variables. The genomic inflation factors (λ) generated by an analysis adding in each of the first 3 PCs, one at a time all remained <1.05 (Table 3-3), indicating minimal population structure and cryptic relatedness in the CGC. PCs were therefore not included as covariates in the linear mixed-effects and ANOVA models.

Table 3-3. Genomic Inflation Factors Generated By a Regression Model Including Varying Numbers of First 3 Principal Components

PC(s) Used	λ
None	1.028
First PC	1.029
First and second PCs	1.030
First, second and third PCs	1.037

PC = principal component; λ = genomic inflation factor.

3.2.12. Significance Thresholds

For individual SNP, *HLA-DRB1* alleles and amino acid association tests, Bonferroni corrected *P*-value thresholds of 0.0007 (69 markers), 0.0029 (17 markers) and 0.0019 (27 markers), respectively were considered significant. For all other tests *P*-values <0.05 were considered significant.

3.2.13. Power Calculations

Power calculations were performed using the Genetic Power Calculator (Purcell, et al., 2003). Five hundred and twenty four patients provided 82.1% power to detect a genetic variant accounting for just 2% of the variance in radiological progression at a Bonferroni corrected *P*-value of 0.0007 (corrected for testing 69 SNPs). The study was, therefore, adequately powered to detect variants of a relatively small effect size.

3.2.14. Statistical Programs

Analyses were performed in R version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria), PLINK version 1.07 (Purcell, et al., 2007) and EIGENSOFT version 5.0.1 (Price, et al., 2006).

3.3. RESULTS

3.3.1. Subjects

524 patients were studied (Table 3-4); most were female (69%) and seropositive (67% RF-positive). Their mean age was 54.7 years. Disease duration was short (median 1.0 month) and disease activity high (mean DAS28 5.88) at study baseline, reflecting the inclusion criteria of the RCTs.

Table 3-4. CARDERA Genetics Cohort Patient Baseline Data

Characteristic		Summary Statistic
Demographic	Number (%) Female	359 (69%)
	Mean Age in Years (95% CI)	54.7 (53.6-55.8)
RA Specific	Median RA Duration in Months (IQR)	1.0 (0.0-4.0)
	Number (%) ACPA-Positive[A]	350 (70%)
	Number (%) RF-Positive	352 (67%)
	Median Larsen Score (IQR)	6.50 (2.00-16.50)
	Mean DAS28 (95% CI)	5.88 (5.77-5.99)
	Mean HAQ (95% CI)	1.56 (1.50-1.62)
Treatment	Number (%) Receiving MTX	161 (31%)
	Number (%) Receiving MTX and CIC	108 (21%)
	Number (%) Receiving MTX and Pred	102 (19%)
	Number (%) Receiving Triple Therapy	107 (20%)
	Number (%) Receiving MTX and Anakinra	46 (9%)

A = ACPA-status missing in 25 patients; CI = confidence interval; IQR = interquartile range; MTX = methotrexate; CIC = ciclosporin; Pred = prednisolone; triple therapy = methotrexate, ciclosporin and prednisolone.

3.3.2. Relationship between Clinical Variables and Larsen Scores

In a linear mixed-effects model incorporating clinical variables only time, age, disease duration and RF significantly associated with Larsen scores; treatment intensity did not significantly associate with Larsen scores (Table 3-5).

Table 3-5. Clinical Factors Associated With Larsen Scores in the CGC

Variable		β (95% CI)	*P*-Value
Time		1.20 (1.17-1.24)	4.987×10^{-36}
Treatment	*MTX*	Reference	-
	MTX/CIC	1.22 (0.95-1.56)	0.115
	MTX/Pred	0.95 (0.74-1.23)	0.708
	Triple Therapy	0.85 (0.67-1.09)	0.200
	Anakinra	1.40 (1.01-1.95)	0.045
Age		1.04 (1.03-1.05)	5.224×10^{-25}
Disease Duration		1.04 (1.02-1.06)	4.230×10^{-05}
RF		1.35 (1.12-1.62)	0.0016

Effect sizes and P-values from a linear mixed-effects model including time and clinical factors only as fixed-effects predictor variables; CI = confidence interval; MTX = methotrexate; CIC = ciclosporin; Pred = prednisolone; triple therapy = methotrexate, ciclosporin and prednisolone.

3.3.3. Radiological Progression Rates

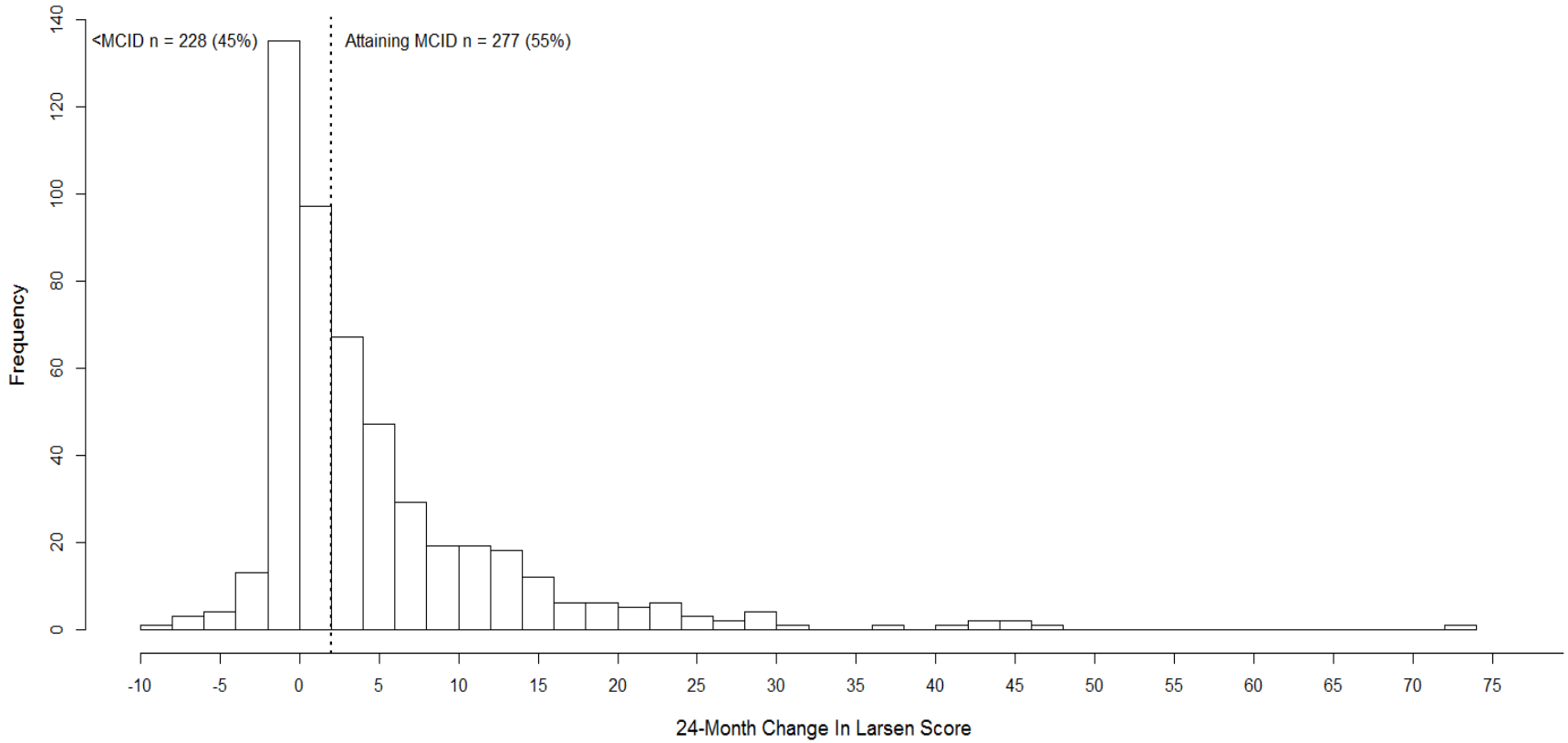
3.3.3.1.In the CGC

Over 2 years, of the 505 patients in the CGC with 24-month X-ray data, 349 (69%) had an increase in their Larsen scores (Figure 3-6). In 277 patients (55%) this increase was above the MCID, defined in early active RA as an increase of ≥2.0 units (Bruynesteyn, et al., 2002).

3.3.3.2.Compared with Observational Studies

Studies rarely report overall X-ray progression rates. Only one observational study has reported the proportion of early RA patients with radiological progression over 2-years (Sanmarti, et al., 2007); 32% of 105 early RA patients receiving sequential DMARD monotherapy and low dose corticosteroids had X-ray progression defined as an increase in Larsen scores of ≥4 units (Sanmarti et al, 2007). Using this definition in the CGC, 199 patients (39%) had progressed; radiological progression rates were, therefore, in line with non-clinical trial cohorts. These findings support the use of the CGC to identify genetic predictors of radiological progression.

Figure 3-6. Histogram of 24-Month Changes in Larsen Scores in the CARDERA Genetics Cohort



MCID = minimal clinically important difference; dotted line represents the MCID in early active RA of 2 units (Bruynesteyn et al, 2002).

3.3.4. Allele/Amino Acid Frequencies

Risk variant frequencies in the CGC and reference meta-analyses were similar (Tables 3-6 to 3-8). Exceptions occurred at the *HLA-DRB1* SNP, rs660895 (meta-analysis frequency 0.46/CGC 0.37), *04:01 (0.31 meta-analysis/0.22 CGC) and HLA-DRβ1 amino acids, histidine at position 13 (His13; 0.45 meta-analysis/0.35 CGC) and valine at position 11 (Val 11; 0.47 meta-analysis/0.36 CGC). These arose because the meta-analysis evaluated ACPA-positive RA (Raychaudhuri et al, 2012); they reduced when only ACPA-positive CGC patients were assessed (frequencies for rs660895, *04:01, His13 and Val11 comprised 0.42, 0.26, 0.40 and 0.41, respectively).

3.3.5. SNP Testing

Only one SNP, rs660895 (*P*=0.0003), had a *P*-value for the genotype*time interaction term that passed the pre-defined significance threshold when tested using the linear mixed-effects model (Table 3-6). This SNP tags *HLA-DRB1**04:01; its β-value of 1.074 (95% CI 1.033-1.117) indicated a 1.07-fold greater annual increase in modified Larsen scores for each copy of the risk (G) allele carried, relative to the reference (A) allele. Findings were similar in a model including ACPA as a covariate in place of RF; only rs660895 (*P*=0.0003) passed multiple testing correction thresholds. Restricting analyses to ACPA-positive patients (removing serology as a covariate) provided a similar β-value for this SNP (β=1.074; 95% CI 1.023-1.126). Although a lower β-value was observed in ACPA-negative patients (β=1.028; 95% CI 0.955-1.106) its upper 95% CI overlapped with the lower 95% CI obtained in the analysis of all RA patients. These findings suggested that the impact of rs660895 on radiological progression was probably independent of ACPA. Figure 3-7 shows the effect of the rs660895 G allele on radiological progression. Results for all 69 tested SNPs ordered by significance are given in Table 3-6.

When using the ANOVA model to test the association between the 69 susceptibility SNPs and the ΔLarsen score the two most strongly associated SNPs − rs660895 (F-statistic 13.70; *P*=0.0002) and rs10175798 (F-statistic 6.08; *P*=0.0140) − were the same as those identified by the linear mixed-effects model. Only rs660895 was significant after correcting for multiple testing.

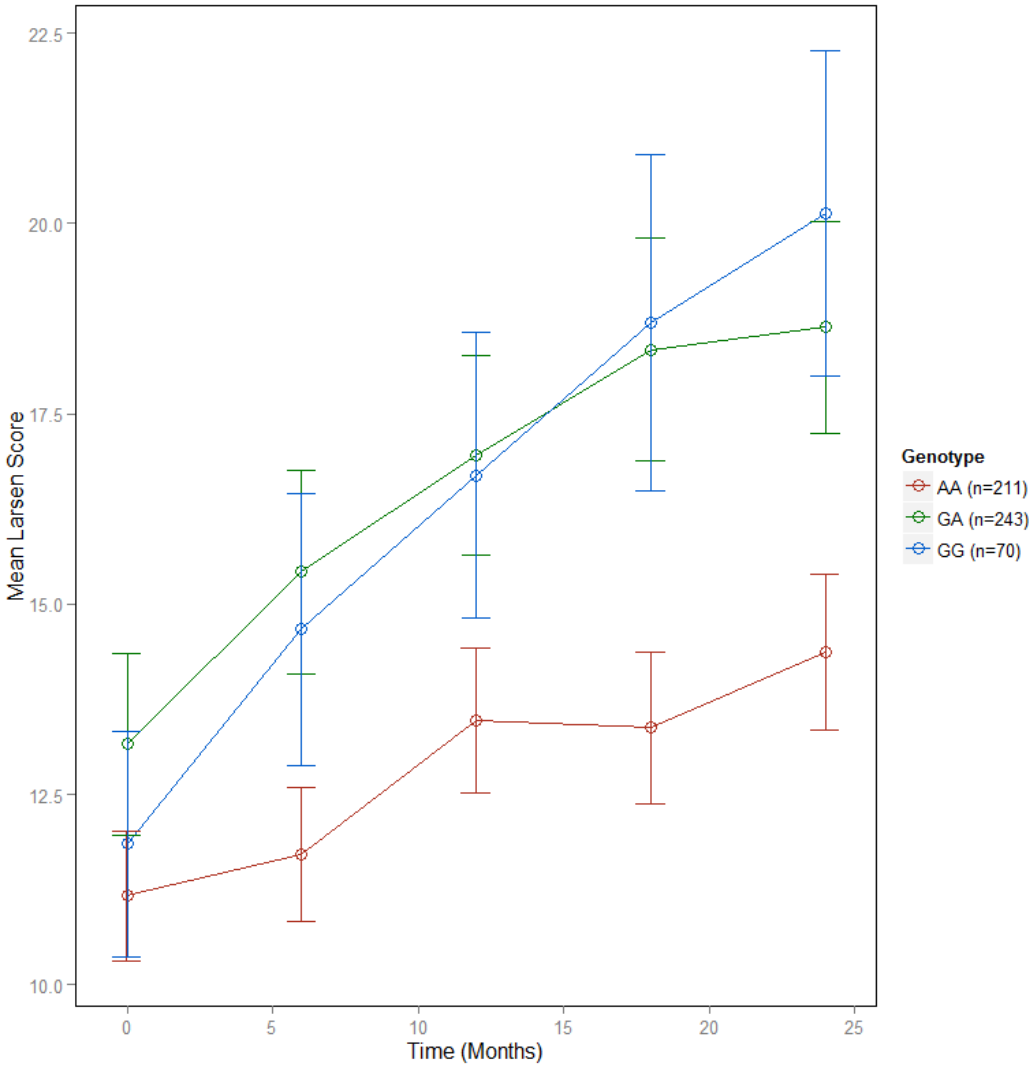
Table 3-6. Association between 69 RA Susceptibility SNPs and Radiological Progression in the CGC Using a Linear Mixed-Effects Model

SNP	Gene	Chr	A1/A2	A1 Frequency		β	*P*-Value
				Meta-Analysis	CGC		
rs660895	*HLA-DRB1*	6	G/A	0.46[A]	0.37	1.074	0.0003
rs10175798	*LBH*	2	A/G	0.62	0.66	1.057	0.0074
rs4452313	*PLCL2*	3	T/A	0.30	0.28	1.053	0.0194
chr17:38031857	*IKZF3-CSF3*	17	G/T	0.48	0.50	0.958	0.0253
rs10774624	*SH2B3-PTPN11*	12	G/A	0.48	0.49	1.044	0.0338
rs3824660	*GATA3*	10	C/T	0.41	0.36	1.039	0.0615
rs71508903	*ARID5B*	10	T/C	0.24	0.20	0.959	0.0836
rs7752903	*TNFAIP3*	6	G/T	0.02	0.04	0.919	0.0950
rs624988	*CD2*	1	T/C	0.40	0.40	0.967	0.0971
rs678347	*GRHL2*	8	G/A	0.23	0.28	0.963	0.0981
rs5987194[B]	*IRAK1*	X	C/G	0.15	0.15	1.048	0.1309
rs998731	*TPD52*	8	T/C	0.49	0.48	0.972	0.1324
rs11889341	*STAT4*	2	T/C	0.23	0.25	1.036	0.1398
rs4239702	*CD40*	20	C/T	0.72	0.74	1.031	0.1576
rs968567	*FADS1-FADS2-FADS3*	11	C/T	0.83	0.84	1.037	0.1657
rs8083786	*PTPN2*	18	G/A	0.15	0.17	0.965	0.1682
rs2561477	*C5orf30*	5	G/A	0.68	0.69	1.028	0.1849
rs73081554	*DNASE1L3-ABHD6-PXK*	3	T/C	0.06	0.09	0.957	0.1937
rs17264332	*TNFAIP3*	6	G/A	0.17	0.26	0.974	0.2239
rs73194058	*IFNGR2*	21	C/A	0.85	0.87	1.036	0.2244
rs5987194[C]	*IRAK1*	X	C/G	0.15	0.15	0.956	0.2284
rs9372120	*ATG5*	6	G/T	0.19	0.21	0.973	0.2381
rs1858037	*SPRED2*	2	T/A	0.68	0.68	0.980	0.3253
chr1:2523811	*TNFRSF14-MMEL1*	1	G/A	0.71	0.67	1.020	0.3354
rs508970	*CD5*	11	A/G	0.45	0.46	1.017	0.4006
rs1893592	*UBASH3A*	21	A/C	0.72	0.73	0.983	0.4171
rs3087243	*CTLA4*	2	G/A	0.55	0.55	1.015	0.4240
rs45475795	*IL2-IL21*	4	A/G	0.08	0.08	0.975	0.4468
rs8026898	*LOC145837*	15	A/G	0.29	0.30	1.017	0.4492
rs1571878	*CCR6*	6	C/T	0.42	0.47	0.986	0.4688
rs10985070	*TRAF1-C5*	9	C/A	0.43	0.41	1.012	0.5212
rs72717009	*FCGR2A*	1	T/C	0.13	0.09	0.980	0.5638
rs2234067	*ETV7*	6	C/A	0.88	0.88	0.983	0.5810

rs34695944	*REL*	2	C/T	0.35	0.39	1.010	0.6105
rs9603616	*COG6*	13	C/T	0.67	0.69	1.011	0.6180
rs11574914	*CCL19-CCL21*	9	A/G	0.32	0.37	1.010	0.6298
rs2301888	*PADI4*	1	G/A	0.65	0.66	0.991	0.6404
rs12140275	*LOC339442*	1	A/T	0.78	0.76	0.990	0.6522
rs28411352	*MTF1-INPP5B*	1	T/C	0.26	0.30	1.009	0.6522
rs2451258	*TAGAP*	6	T/C	0.64	0.62	0.991	0.6575
rs2228145	*IL6R*	1	A/C	0.64	0.60	0.991	0.6624
rs13330176	*IRF8*	16	A/T	0.24	0.27	1.009	0.6717
rs909685	*SYNGR1*	22	A/T	0.31	0.32	1.008	0.6926
rs3218251	*IL2RB*	22	A/T	0.27	0.28	0.991	0.6931
rs17668708	*PTPRC*	1	C/T	0.89	0.90	1.012	0.7098
rs331463	*TRAF6-RAG1/2*	11	T/A	0.87	0.87	1.011	0.7108
rs1633360	*CDK4*	12	T/C	0.59	0.60	0.993	0.7338
rs947474	*PRKCQ*	10	A/G	0.82	0.83	0.992	0.7412
chr21:35928240	*RCAN1*	21	C/T	0.86	0.89	1.010	0.7490
rs8133843	*RUNX1-LOC100506403*	21	A/G	0.63	0.65	1.006	0.7559
rs10790268	*CXCR5*	11	G/A	0.81	0.83	0.993	0.7656
rs67250450	*JAZF1*	7	T/C	0.79	0.81	0.993	0.7791
rs773125	*CDK2*	12	A/G	0.62	0.62	1.005	0.7871
rs1950897	*RAD51B*	14	T/C	0.67	0.73	0.994	0.7890
rs3806624	*EOMES*	3	G/A	0.47	0.44	0.995	0.7896
rs9826828	*IL20RB*	3	A/G	0.02	0.01	1.021	0.8146
rs9653442	*AFF3*	2	C/T	0.46	0.51	1.004	0.8189
rs706778	*IL2RA*	10	T/C	0.42	0.43	0.996	0.8353
rs6715284	*CFLAR-CASP8*	2	G/C	0.11	0.11	1.005	0.8745
rs8032939	*RASGRP1*	15	C/T	0.25	0.26	1.003	0.8830
rs11933540	*C4orf52*	4	C/T	0.31	0.34	0.997	0.8874
rs4409785	*CEP57*	11	C/T	0.19	0.19	0.997	0.8912
rs2736337	*BLK*	8	C/T	0.25	0.25	0.997	0.8940
rs4272	*CDK6*	7	G/A	0.23	0.23	1.003	0.8946
rs1980422	*CD28*	2	C/T	0.26	0.22	1.002	0.9251
rs2476601	*PTPN22*	1	A/G	0.09	0.15	1.002	0.9284
rs2105325	*LOC100506023*	1	C/A	0.76	0.76	0.999	0.9664
chr7:128580042	*IRF5*	7	G/A	0.46	0.51	1.001	0.9688
rs34536443	*TYK2*	19	G/C	0.97	0.97	1.002	0.9716
rs1516971	*PVT1*	8	T/C	0.90	0.88	1.001	0.9723

Alleles ordered by significance (most significant listed first); A = A1 frequency data from Raychaudhuri et al meta-analysis (data unavailable in Eyre et al meta-analysis); B = result in females; C = result in males; Chr = chromosome; unless otherwise stated reference meta-analysis is by Okada et al (Okada et al, 2013).

Figure 3-7. Mean Larsen Scores Stratified By rs660895 (*04:01) Genotype



Mean Larsen scores with standard error bars shown for each time point.

3.3.6. *HLA-DRB1* Allele Testing

Only the *HLA-DRB1**04:01 (*P*=0.0006) allele had a *P*-value for the genotype*time interaction term that passed the pre-defined significance threshold when tested using the linear mixed-effects model (Table 3-7). Its β-value of 1.087 (95% CI 1.037-1.139) indicated a 1.09-fold greater annual increase in modified Larsen scores for each copy of the *04:01 allele carried, relative to a non-carrier. Findings were similar in a model including ACPA as a covariate in place of RF, with only *04:01 having a significant association (*P*=0.0020). Restricting analyses to ACPA-positive patients revealed a similar β-value (β=1.068; 95% CI 1.010-1.130). On restricting the analysis to ACPA-negative patients (β=1.033; 95% CI 0.931-1.147) the β-value

upper 95% CI overlapped with the lower 95% CI obtained in the analysis of all cases. These findings suggested its effect was independent of ACPA. Results for all 17 tested *HLA-DRB1* alleles ordered by significance are given in Table 3-7.

When using the ANOVA model to test the association between *HLA-DRB1* susceptibility alleles and the ΔLarsen score, only *04:01 had a significant association (F-statistic 10.08; *P*=0.0016).

Table 3-7. Association between RA *HLA-DRB1* Susceptibility Alleles and Radiological Progression in the CGC Using a Linear Mixed-Effects Model

***HLA-DRB1* Allele**	**Allele Frequency**		**β**	***P*-Value**
	Meta-Analysis	**CGC**		
*04:01	0.309	0.219	1.087	0.0006
*04:08	0.017	0.011	1.238	0.0180
*15:01	0.089	0.103	0.948	0.1031
*13:02	0.012	0.021	0.893	0.1078
*03:01	0.082	0.111	0.950	0.1086
*13:01	0.021	0.017	0.905	0.1672
*11:03	0.002	0.010	1.110	0.2890
*07:01	0.064	0.085	0.967	0.3335
*10:01	0.020	0.013	0.922	0.3373
*01:01	0.133	0.136	1.019	0.5078
*14:01	0.011	0.015	1.049	0.5569
*11:01	0.028	0.032	0.978	0.6873
*04:04	0.091	0.092	1.014	0.6906
*14:04	0.001	0.001	1.083	0.7958
*04:05	0.012	0.011	0.981	0.8382
*11:04	0.008	0.018	0.989	0.8768
*08:01	0.009	0.017	1.009	0.9063

Alleles ordered by significance (most significant listed first); reference meta-analysis by Raychaudhuri et al (Raychaudhuri et al, 2012).

3.3.7. HLA Protein Amino Acid Testing

Two amino acid polymorphisms had *P*-values for the genotype*time interaction term that passed the pre-defined significance threshold when tested using the linear mixed-effects model (Table 3-8). These comprised His13 (β=1.073; 95% CI 1.031-1.116; *P*=0.0005) and Val11 (β=1.067; 95% CI 1.026-1.110; *P*=0.0013) in HLA-

DRβ1. Repeating the analysis using ACPA as a covariate in place of RF revealed similar results (His13 *P*=0.0006; Val11 *P*=0.0013). Restricting analyses to ACPA-positive patients revealed similar findings (His 13 β=1.071 (95% CI 1.020-1.125); Val11 β=1.065 (95% CI 1.014-1.119)). Restricting analyses to ACPA-negative patients, revealed β-value upper 95% CIs that overlapped with the lower 95% CIs obtained in the analysis of all cases (His 13 β=1.022 (95% CI 0.950-1.100); Val11 β=1.019 (95% CI 0.946-1.096)). When using the ANOVA model to test the association between HLA susceptibility proteins and the ΔLarsen score, His13 (F-statistic 14.08; *P*=0.0002) and Val11 (F-statistic 14.42; *P*=0.0002) in HLA-DRβ1 remained the most significant markers.

As His13 and Val11 are in tight LD with each other (r^2=0.943) a conditional analysis was undertaken. Conditioning on His13 eliminated the effect of Val 11 (linear mixed-effects model *P*=0.5279; ANOVA model *P*=0.5264). Conditioning on Val11 eliminated the effect of His13 (linear mixed-effects model *P*=0.1498; ANOVA model *P*=0.7946). This indicates the significance at both these positions is driven by their high LD.

None of the tested polymorphisms in the SE region (positions 71 and 74) had *P*-values <0.05. However, the SE itself (sequences QRRAA, RRRAA and QKRAA in positions 70-74 (Gregersen et al, 1987)) significantly associated with radiological progression when the classical SE alleles (*01:01, *01:02, *04:01, *04:04, *04:05, *04:08, *10:01) were grouped together (linear mixed-effects model *P*=0.0001; ANOVA model *P*=0.0001). The β-value of 1.080 (95% CI 1.038-1.123) indicated a 1.08-fold greater annual increase in modified Larsen scores for each copy of the SE carried, relative to a non-carrier. Other rarer SE alleles were not available for analysis in the imputed dataset.

Although His13 and Val11 are not in LD with the individual SE positions tested, there is a marked correlation between their presence and carrying a SE sequence. His13 and Val11 are both encoded by the classical *HLA-DRB1* SE alleles *04:01, *04:04, *04:05, *04:08 and *10:01; they are also encoded by the classical *HLA-DRB1* non-SE alleles *04:02, *04:03 and *04:07, but these are rare with an allele

frequency <0.01 in ACPA-positive RA (Raychaudhuri et al, 2012). A conditional analysis incorporating the SE as a modelling covariate was therefore undertaken. Conditioning on the SE eliminated the effect of His13 (linear mixed-effects model P=0.2412; ANOVA *P*=0.1409) and Val11 (linear mixed-effects model P=0.4453; ANOVA *P*=0.1371) on Larsen score progression. This indicated their association was probably driven by their correlation with the SE.

Table 3-8. Relationship between HLA Amino Acid Polymorphisms and Radiological Progression in the CGC Using a Linear Mixed-Effects Model

HLA Amino Acid	Amino Acid Frequency		β	*P*-Value
	Meta-Analysis	CGC		
DRβ1; position 13; Histidine	0.449	0.345	1.073	0.0005
DRβ1; position 11; Valine	0.470	0.359	1.067	0.0013
DRβ1; position 13; Serine	0.176	0.238	0.956	0.0499
DRβ1; position 71; Lysine	0.397	0.339	1.042	0.0573
DRβ1; position 71; Alanine	0.092	0.109	0.943	0.0689
DRβ1; position 74; Alanine	0.801	0.744	1.040	0.0809
DRβ1; position 71; Glutamic Acid	0.052	0.065	0.935	0.0855
DRβ1; position 11; Serine	0.205	0.274	0.965	0.0926
DRβ1; position 74; Arginine	0.082	0.111	0.950	0.1086
DRβ1; position 11; Proline	0.104	0.120	0.952	0.1152
DRβ1; position 13; Arginine	0.104	0.120	0.952	0.1152
DRβ1; position 11; Aspartic Acid	0.013	0.014	0.896	0.2024
DPβ1; position 9; Histidine	0.035	0.043	0.954	0.3203
DRβ1; position 11; Glycine	0.064	0.085	0.967	0.3335
DRβ1; position 13; Tyrosine	0.064	0.085	0.967	0.3335
DRβ1; position 74; Glutamine	0.064	0.085	0.967	0.3335
DPβ1; position 9; Phenylalanine	0.799	0.733	1.014	0.5204
DRβ1; position 74; Leucine	0.013	0.021	1.037	0.5872
DRβ1; position 11; Leucine	0.145	0.148	1.011	0.6949
B; position 9; Tyrosine	0.643	0.655	1.008	0.7095
DRβ1; position 74; Glutamic Acid	0.039	0.039	0.982	0.7182
DRβ1; position 13; Glycine	0.028	0.036	1.018	0.7257
B; position 9; Histidine	0.227	0.218	0.993	0.7602
DRβ1; position 13; Phenylalanine	0.178	0.176	0.992	0.7612
DRβ1; position 71; Arginine	0.458	0.488	1.005	0.8139
DPβ1; position 9; Tyrosine	0.165	0.224	0.995	0.8445
B; position 9; Aspartic Acid	0.130	0.128	0.996	0.8949

Amino acids ordered by significance (most significant listed first); reference meta-analysis by Raychaudhuri et al (Raychaudhuri et al, 2012).

3.3.8. Relationship between wGRS and Radiological Progression

A significant association was observed between a full wGRS and radiological progression when assessed using the linear mixed-effects (*P*=0.0177; β=1.038; 95% CI =1.007-1.071) and ANOVA (F-statistic=6.315; *P*-value=0.0123) models. No association was observed between a non-HLA wGRS score and radiological progression when assessed using the linear mixed-effects (*P*=0.9381; β=1.002; 95% CI=0.961-1.044) and ANOVA (F-statistic=0.035; *P*-value=0.8509) models.

3.4. DISCUSSION

This study has confirmed the influence of HLA susceptibility variants on radiological progression in RA, by demonstrating a significant association between the main HLA risk variants (the *HLA-DRB1**04:01 allele and amino acid polymorphisms valine at position 11 and histidine at position 13 in the HLA-DRβ1 molecule) and Larsen score progression in early, active RA patients. This effect appears to be independent of ACPA status, with similar effect sizes observed on restricting the analysis to ACPA-positive individuals.

This study has also demonstrated that no significant association exists between non-HLA susceptibility loci and radiological progression in early, active RA patients, when evaluating these variants both individually and cumulatively, using a wGRS. Although the sample size of 524 patients is substantially smaller than the combined GWASs evaluated to identify these susceptibility loci (Okada et al, 2013), the CGC was adequately powered (at >80%) to detect genetic markers accounting for just 2% of the variance in X-ray progression. Furthermore, this study was able to identify known predictors of joint destruction including RF, ACPA and the SE (Van Der Woude, et al., 2010b, Gorman et al, 2004). This indicates that non-HLA susceptibility loci do not have a clinically relevant association with radiological progression in early, active RA patients. It also suggests that the non-HLA genetic architectures of RA susceptibility and severity probably, at least in part, differ.

The impact of *HLA-DRB1**04:01 on radiological damage is well described. Gorman *et al* reported a significant relationship between *04:01 and erosions in a meta-analysis of 466 RA patients: OR for erosions with 1 allele copy 3.1 (95% CI 2.0–5.0)

(Gorman et al, 2004). Mewar *et al* reported a significant association (*P*=0.0059) between S2 allele (*04:01 and *13:03) carriage and higher Larsen scores in 962 RA patients (Mewar et al, 2008). In contrast the impact of HLA-DRβ1 molecule polymorphisms on X-ray damage has not previously been evaluated. In this study the most significant polymorphisms (Val11 and His13) for RA development (Raychaudhuri et al, 2012) were also the only ones to influence radiological progression. Both are located in peptide-binding grooves; they could therefore mediate articular damage through influencing arthritogenic peptide presentation within the joint. Conditional analyses indicated the association at both positions was driven by the tight LD between them, although it was not possible to establish which had the dominant association. Furthermore, conditional analyses also suggested their association was driven by a high correlation with the SE at positions 70-74. Due to the broad LD across the MHC region (De Bakker, et al., 2006) alongside the modest size of the CGC it was not possible to characterise which HLA amino acid positions had a SE-independent association with joint destruction. Further work is required to assess this on an HLA-wide scale across multiple large cohorts, in a similar manner as undertaken for RA susceptibility (Raychaudhuri et al, 2012) .

Several studies have reported significant associations between susceptibility loci and radiological damage in RA. Recent examples include *TRAF1* (Viatte, et al., 2013), *IL2RA* (Knevel et al, 2013a) and *C5orf30* (Teare et al, 2013); none of these associated with radiological progression in the CGC (*P*-values for the linear mixed-effects model were 0.5212, 0.8353 and 0.1849, respectively). As the published reports used two SNPs that were not in LD with those tested in this study, the precise SNPs were tested separately (Table 3-9); this confirmed no association. The two most likely explanations for this non-replication are firstly, that the CGC population − which comprises patients with early, severely active disease − is inherently different to the populations evaluated in these observational studies, which assessed patients with a range of disease durations and severities and secondly, the CGC looked at predictors of X-ray progression in the first 2 years of disease and other studies evaluated damage over longer time periods. Other possible explanations are that previous studies were not able to fully adjust for the effects of treatment, which may be an important confounding variable; some studies assessed genotype effects

on the severity of X-ray damage, as opposed to its impact on progression over time; and the use of different radiological scoring systems across studies. The lack of replication in the CGC indicates that published predictors of radiological damage are not generalisable across all RA patient populations.

Table 3-9. Testing Susceptibility SNPs Previously Associated With Radiological Damage in the CGC

SNP	Candidate Gene	Chr	A1/A2	A1 Freq		β	*P*-Value
				1,000 Genomes*	CGC		
rs26232	*C5orf30*	5	A/G	0.28	0.31	0.975	0.2550
rs2900180	*TRAF1/C5*	9	A/G	0.33	0.33	1.003	0.8914
rs2104286	*IL2RA*	10	G/A	0.14	0.27	1.003	0.8926

*Chr = chromosome; P-values from linear mixed-effects model (including RF as a covariate), *1,000 Genomes EUR population panel.*

This study has several strengths. As the CGC represents a clinical trial cohort, all patient assessments were performed in a timely, highly standardised manner. The inclusion of individuals with severely active disease meant any findings were directly relevant to patient populations upon which current NICE guidelines for RA management have been based; it also ensured that the effect of baseline disease activity as a confounding factor was accounted for. The regularity of Larsen scoring optimised the power to detect longitudinal associations. Additionally, the randomisation to treatment groups ensured that treatment effects could be fully adjusted for. Finally, it was able to identify known predictors of X-ray damage, spanning clinical (disease duration), serological (RF and ACPA) and genetic (HLA variants) factors; this indicates the study was adequately powered to detect clinically relevant predictors of Larsen scores.

This study also has a number of important limitations. Firstly, its size of 524 patients, which is modest for a genetic study, limited its power to detect variants of a small effect size. However, these variants would be of too small an effect size to influence clinical decision-making. Secondly, Larsen scores were only evaluated over 24-months; genetic variants could exert their effects over longer time periods. Thirdly, outcome data were missing in some patients. However, the linear mixed-effects

model was able to accommodate missing data, by making the assumption that these were MAR (Higgins and Green, 2011); as previously discussed this assumption appeared reasonable with most missing data occurring in CARDERA-2 patients at 6 and 18 months as it was not study protocol for X-rays to be performed at these time points. Additionally, similar results were obtained for the ANOVA model, which excluded individuals with missing Larsen score data. Fourthly, as the CGC comprised individuals from RCTs, radiological progression rates could be lower than those seen in observational studies, further reducing the analysis power. This appears unlikely with radiological progression levels similar to an observational study of early RA patients (Sanmarti et al, 2007). Finally, the linear radiological progression rates observed in the CGC are unlikely to persist over time. Therefore, the β-values associated with each variant may only be applicable to the first few years of disease.

Further work is required to define the genetic basis of radiological progression in RA. This study suggests that a candidate gene approach testing variants known to associate with RA development is unlikely to adequately characterise this trait. Any non-HLA variants evaluated in this study, if significant, were of too small an effect size for an association to be detected. A genome-wide approach may be more effective. Although such an analysis would result in a loss of power, by testing markers across the breadth of the genome any loci not previously considered *a priori* to have an impact on X-ray damage may be detected, provided they are of a large enough effect size.

In conclusion this study has demonstrated an association between HLA susceptibility variants and radiological progression in a unique cohort of early, active RA patients assessed in a clinical trial setting. No significant relationship was seen between non-HLA susceptibility variants and radiological progression. This suggests that these markers do not have a clinically relevant association with X-ray progression in early, active RA patients; it raises the possibility that the non-HLA genetic architectures of RA susceptibility and severity probably, at least in part, differ. HLA-wide and genome-wide analyses are required to better characterise the genetic architecture of radiological progression in RA.

CHAPTER 4. GENETIC PREDICTORS OF X-RAY PROGRESSION

4.1. BACKGROUND

Most studies have used a candidate gene design – in which a single polymorphism or set of polymorphisms near a single gene are evaluated (Lunetta, 2008) – to identify loci that associate with radiological damage in RA. This approach, undertaken due to the limited size of most study cohorts, has several limitations. Firstly, it only assesses pre-specified markers that are considered likely to influence radiological damage. As the genome comprises over 3 billion base pairs, this approach is highly likely to miss important loci. Secondly, the studies that use it often fail to correct for population stratification (Daly and Day, 2001). Radiological damage may be more likely to occur in a subpopulation of a different ancestry, which may also harbour different allele frequencies of the candidate marker; this increases the likelihood of obtaining false-positive results. Thirdly, the main candidate genes for RA X-ray progression are those associated with its susceptibility; as demonstrated in the previous chapter of this thesis the effect of non-HLA RA susceptibility SNPs on joint destruction is uncertain with no association observed in 524 patients with early, active disease. Taken together these factors suggest a genome-wide approach is required to better characterise the genetic architecture of joint destruction in RA.

To date only two GWASs have been undertaken that have examined genetic associations with radiological damage in RA. The first study, by Knevel *et al* identified a significant, replicable association between SNPs in the *SPAG16* locus and radiological progression in 384 ACPA-positive established RA patients (Knevel et al, 2013b). The second study, by De Rooy *et al* identified two SNPs in the *RAD51L1/ZFP36L1* and *MMP9* loci that had replicable associations with radiological damage in an analysis of 646 early RA patients (De Rooy et al, 2013b). As the latter study utilised the ImmunoChip array it assessed loci associated with immune-mediated diseases on a genome-wide scale, as opposed to evaluating all genomic regions.

The aim of this study was to undertake a genome-wide analysis to identify novel genetic associations with radiological progression in early, active RA patients enrolled to the CGC. Genetic markers across the breadth of the genome with established links to immune-mediated diseases present on the ImmunoChip were

tested for their association with modified Larsen score progression over two years using a linear mixed-effects model.

4.2. PATIENTS AND METHODS

4.2.1. Subjects

The CGC (ethical approval, inclusion and exclusion criteria and RCT protocols) has been described in detail in Chapter 3 of this thesis. In brief, it comprises 524 patients with early, active RA previously enrolled to two RCTs evaluating intensive combination treatment and anakinra. Patients had modified Larsen scores recorded every 6-12 months for 2 years.

4.2.2. Genotyping

Patients were genotyped on the ImmunoChip. Post-QC 138,873 genetic markers were available for analysis (genotyping rate 0.999; full QC details provided in Chapter 3). This current study evaluated the 138,488 genetic markers on chromosomes 1-22, and the X-chromosome.

4.2.3. Linear Mixed-Effects Model

The association between Larsen score progression and each genetic marker was examined using the linear mixed-effects model described in Chapter 3. Having shown the validity of this approach compared with an ANOVA model evaluating the ΔLarsen, a prospective decision was made to use the linear modelling only, as the use of repeated measures optimised the power of the study to detect significant associations (Guo et al, 2013).

Log_e (Larsen score + 1) was used as the response variable and time (years), clinical variables (treatment, age, disease duration, and RF status), genotype and a genotype*time interaction term, were included as fixed-effects predictor variables. All β-estimates were back-transformed to the original Larsen scale. The genotype*time term β provided information on the annual increase in Larsen score per risk allele copy relative to the reference genotype; *P*-values for this term

provided information on the variant's role in radiological progression. Examination of the model residuals confirmed a good fit of the model.

Testing of genetic markers on the X-chromosome was undertaken separately in males and females. *P*-values from both sexes for each marker were subsequently combined to give a single *P*-value using Stouffer's method, in which individual *P*-values are transformed into the quantiles of a standard normal; the *P*-value of the average of the quantiles is subsequently identified (Burns, 2004).

4.2.4. Genomic Inflation

The genomic inflation factor (λ), which represents the ratio of the median χ^2-test statistic to the expected median (0.456), was calculated for the genotype*time interaction term across all tested genetic markers (De Bakker, et al., 2008). This provided an estimate of the inflation of false-positive rates due to population structure; a $\lambda \leq 1.1$ is considered an acceptable level of inflation (Yamaguchi-Kabata, et al., 2008).

4.2.5. Significance Thresholds

For the association of genetic markers with radiological progression a pre-defined significance threshold of $P<9.18 \times 10^{-7}$ was used. This was based on a Bonferroni correction for all uncorrelated genetic markers present in the post-QC ImmunoChip dataset (n=54,445). This was based on an $r^2 \geq 0.8$. For all other tests *P*-values <0.05 were considered significant.

4.2.6. Power Calculations

Power calculations were performed using the Genetic Power Calculator (Purcell et al, 2003). Five hundred and twenty four patients provided 83.7%, 50.3% and 12.1% power to detect a genetic variant accounting for 7%, 5% and 3% of the variance in radiological progression, respectively at a Bonferroni corrected *P*-value of 9.18×10^{-7}. The study was therefore only powered to detect variants of a large effect size.

4.2.7. Statistical Programs

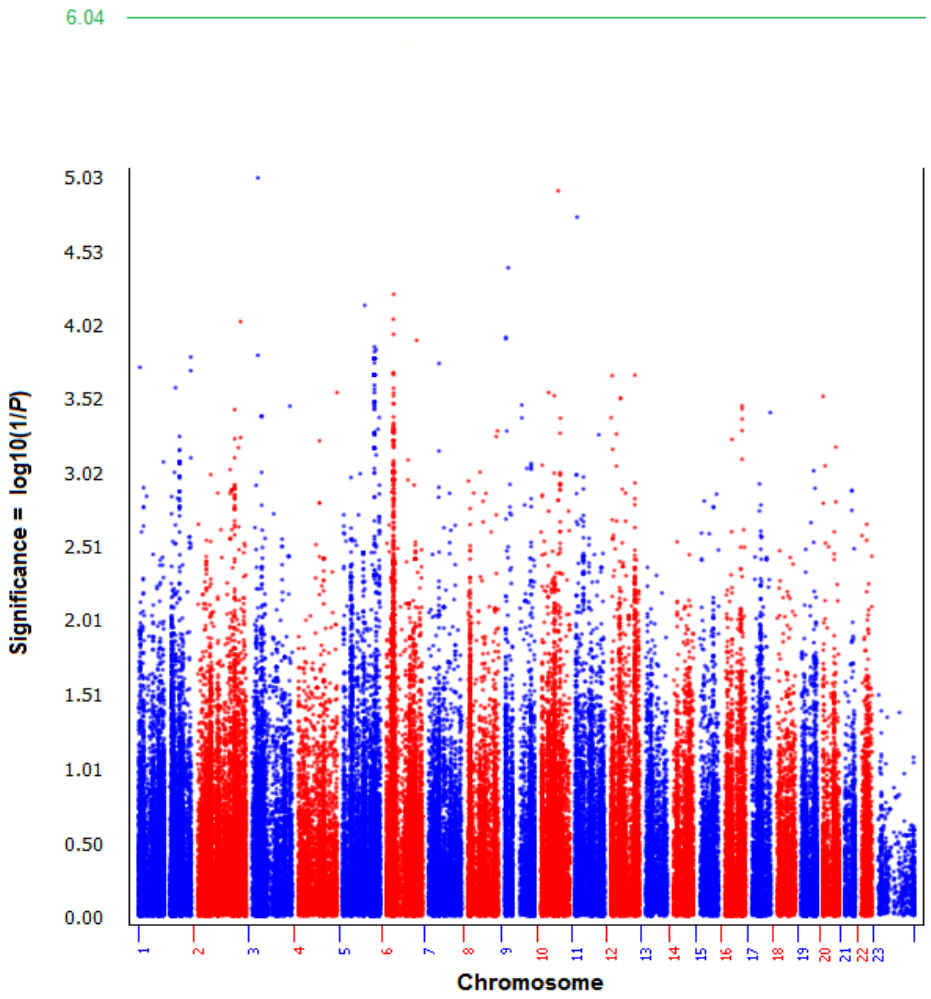
Analyses were performed in R version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria) and PLINK version 1.07 (Purcell et al, 2007).

4.3. RESULTS

4.3.1. ImmunoChip Marker Associations with Larsen Score Progression

None of the 138,488 tested genetic markers had a genome*time interaction term *P*-value that passed the pre-defined significance threshold of $P<9.18x10^{-7}$ (Figure 4-1).

Figure 4-1 Manhattan Plot of Genetic Associations with Radiological Progression in the CARDERA Genetics Cohort

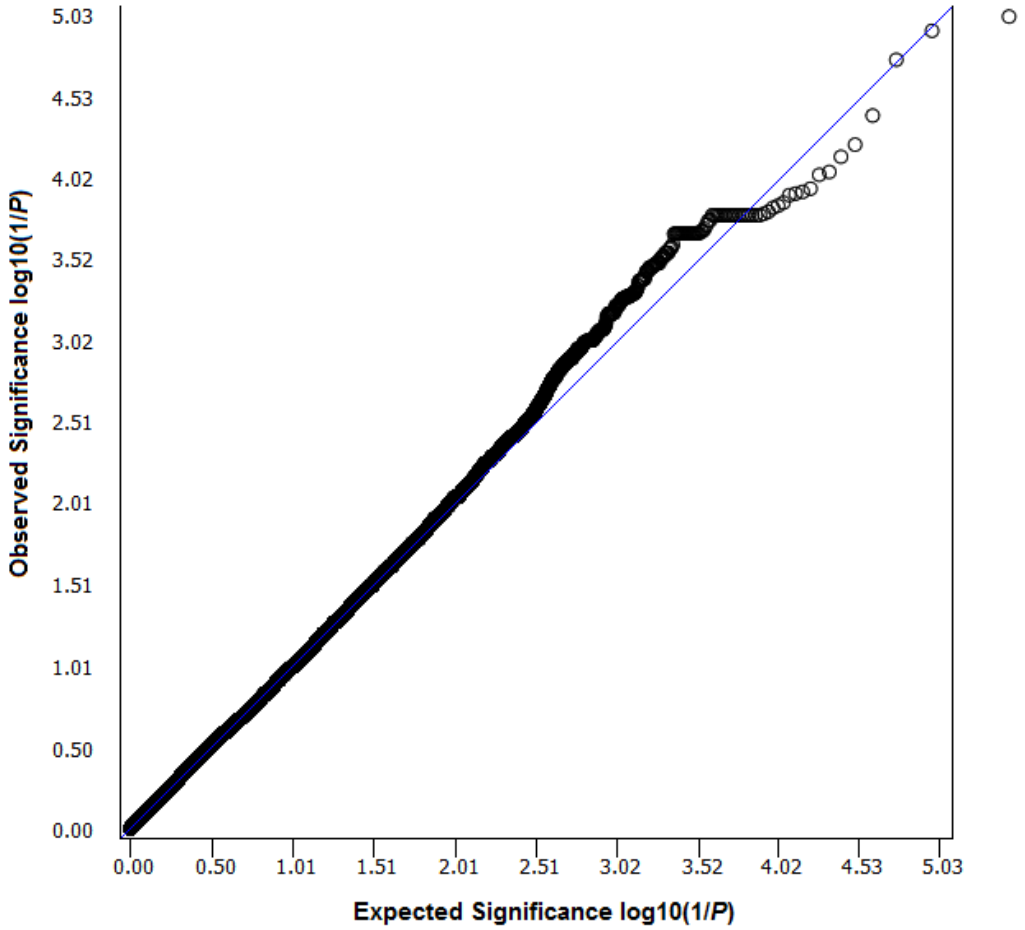


*P-values from genotype*time interaction term; green line represents the significance threshold of $P<9.18x10^{-7}$; Manhattan Plot constructed using SNPEVG version 3.2 (Wang, et al., 2012).*

4.3.2. Genomic Inflation Factor

The λ was 1.052, indicating the deviation of the observed χ^2-test statistic from the expected null distribution was within acceptable limits. This is demonstrated in a Q–Q plot of the observed versus the expected *P*-values for the genotype*time interaction term (Figure 4-2). Genomic control correction was, therefore, not required.

Figure 4-2. Quantile–Quantile Plot of the Genotype*Time *P*-Values



QQ plot constructed using SNPEVG version 3.2 (Wang et al, 2012).

4.3.3. Most Significant Associations with Larsen Score Progression

Eight SNPs had *P*-values $<1x10^{-4}$ for the genotype*time interaction term (Table 4-1). None were in LD. The two most significant SNPs were rs35309890 on chromosome 3 in the *EOMES* locus ($P=9.35x10^{-6}$) and rs12356376 on chromosome 10 in the *ZMIZ1* locus ($P=1.15x10^{-5}$).

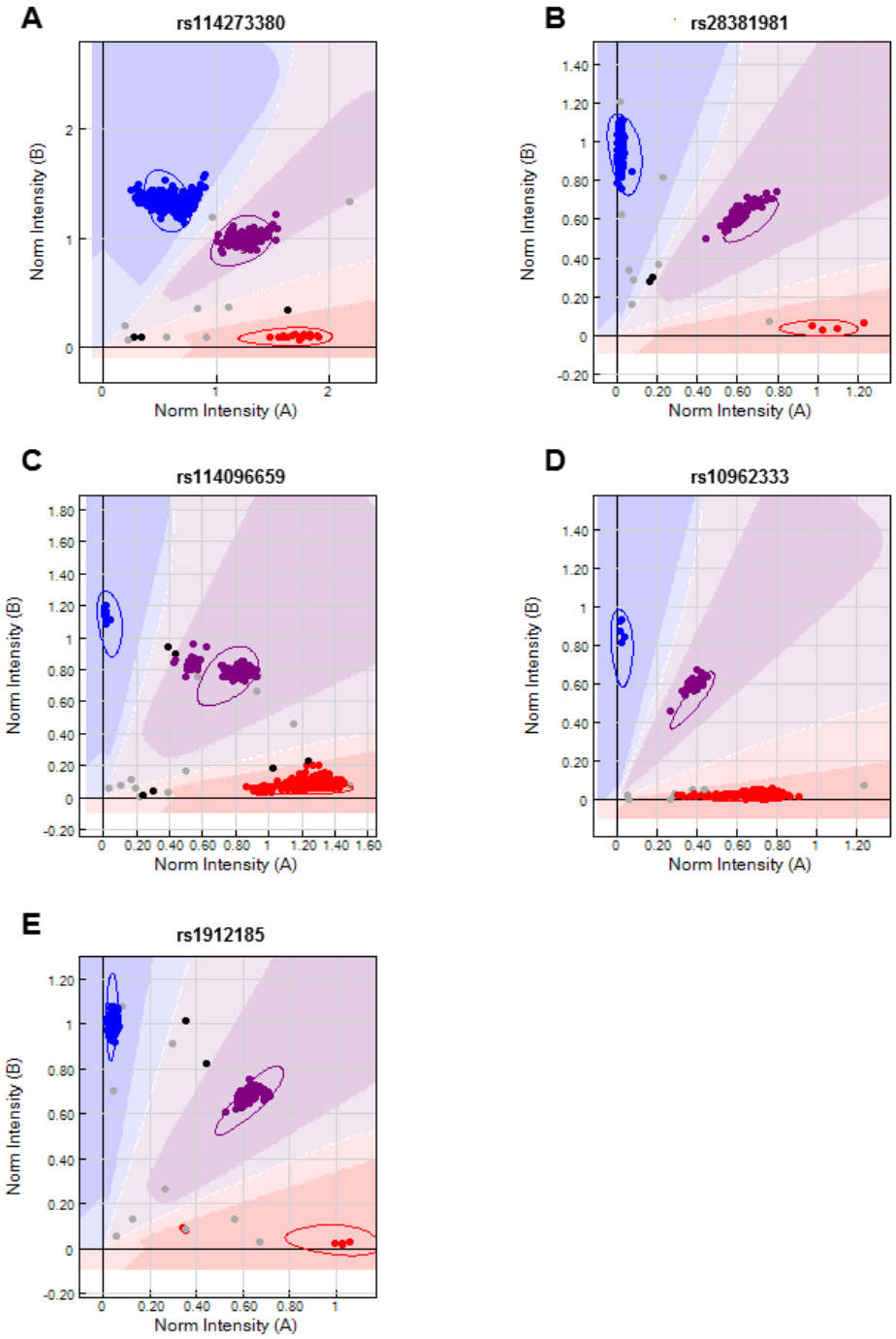
There were some differences in MAFs between the CGC and 1,000 genomes European (EUR) population reference panel (1000 Genomes Project Consortium, et al., 2012). The largest was at the chromosome 6 marker rs114273380, which had a MAF of 0.22 in the CGC and 0.08 in the EUR panel. Other minor differences were seen at rs28381981 and rs114096659, which had slightly higher MAFs in the CGC and rs10962333 and rs1912185, which had slightly lower MAFs in the CGC (differences all ≤ 0.08). These were not due to genotype calling errors, as visual examination of the intensity clustering plots confirmed adequate cluster separation and genotype calling (Figure 4-3). Furthermore all eight SNPs in Table 4-1 were in HWE; only one SNP (rs10962333) had a HWE exact test *P*-value <0.05, although this did not deviate from HWE when a Bonferroni correction was applied ($P=0.0084$). These differences probably represent true allele frequency variations between the CGC (n=524 individuals) and EUR panel (n=379 individuals).

Table 4-1. SNPs with *P*-values <1.00x10^{-04} for An Association with Radiological Progression in the CARDERA Genetics Cohort

SNP	Chr	Base Pair Position*	Loci	A1	MAF		*P*-Value	β (95% CI)
					CGC	1,000 Genomes**		
rs35309890	3	27740229	*EOMES*	A	0.01	0.01	9.35x10^{-06}	1.54 (1.28-1.87)
rs12356376	10	81062107	*ZMIZ1*	A	0.06	0.08	1.15x10^{-05}	1.19 (1.10-1.29)
rs28381981	11	5686266	*TRIM5*	A	0.08	0.03	1.73x10^{-05}	1.16 (1.09-1.25)
rs10962333	9	16210610	*C9orf92*	G	0.05	0.13	3.84x10^{-05}	1.19 (1.10-1.29)
rs114096659	6	32680576	*HLA-DQB1/HLA-DQA2*	C	0.18	0.12	5.81x10^{-05}	1.11 (1.05-1.17)
rs80119111	5	102339176	*PAM*	A	0.02	0.01	6.91x10^{-05}	1.32 (1.15-1.52)
rs114273380	6	31438096	*HCG26*	A	0.22	0.08	8.58x10^{-05}	1.10 (1.05-1.15)
rs1912185	2	215194668	*SPAG16*	A	0.09	0.14	8.93x10^{-05}	1.14 (1.07-1.22)

*Chr = chromosome; A1 = allele 1 (minor allele); MAF = minor allele frequency; P-values from genotype*time interaction term; SNPs ordered by significance; *from GRCh37 assembly; ** MAFs from 1,000 Genomes phase 1 European population panel (CEU/IBS/CBR/FIN/TSI populations) (1000 Genomes Project Consortium et al, 2012)*

Figure 4-3 Intensity Clustering Plots for SNPs with Allele Frequency Differences

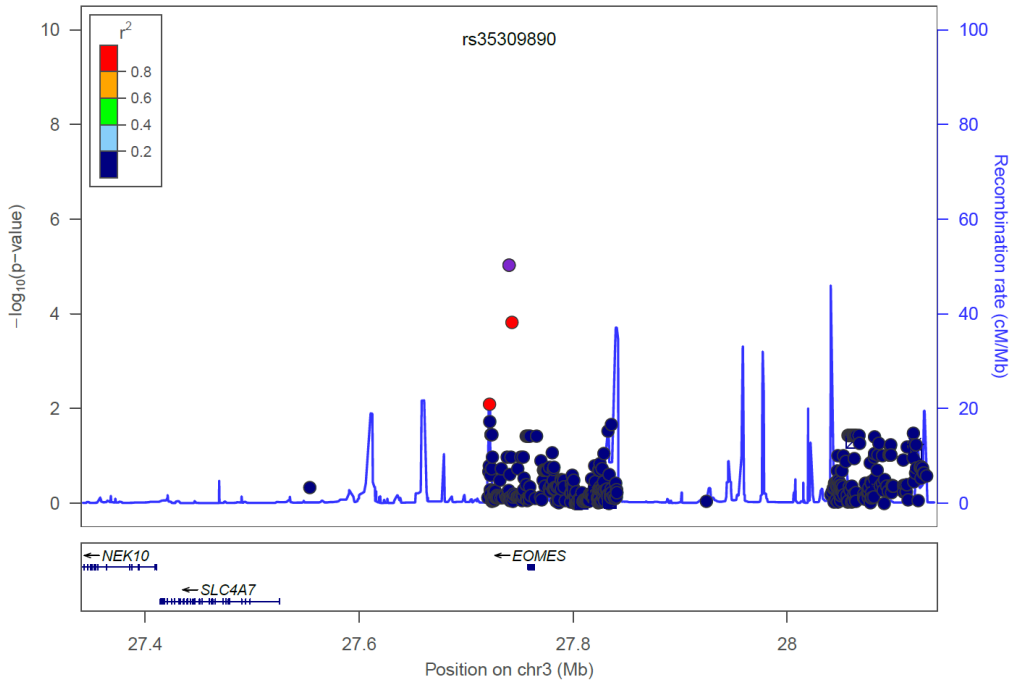


Panel A = rs114273380 (HCG26 locus); Panel B = rs28381981 (TRIM5 locus); Panel C = rs114096659 (HLA-DQB1/HLA-DQA2 locus); Panel D = rs10962333 (C9orf92 locus); Panel E = rs1912185 (SPAG16 locus); blue, purple and red dots represent called genotypes; clustering plots are for all genotyped cases pre-QC procedures.

4.3.3.1.rs35309890

The most significant SNP was rs35309890, which resides on chromosome 3 downstream of the *EOMES* gene. *EOMES* encodes the protein eomesodermin, which is a T-box transcription factor expressed in activated CD8+ T cells; experimental studies suggest it is important in regulating CD8+ T cell activity (Intlekofer, et al., 2005). Figure 4-4 represents a regional plot of the association between the other SNPs in this area and radiological progression. Two SNPs were in LD (R^2>0.8; defined using the 1,000 Genomes EUR population panel) with rs35309890. One of these (rs1112524) had a greater trend towards an association with Larsen score progression (genotype*time P=1.51x10^{-4}) compared to regional SNPs not in LD. Although the other SNP (rs7614831) had a substantially weaker association (P=0.0080), examining LD structure in the CGC revealed only moderate LD (r^2=0.755) between rs35309890 and rs7614831 in this dataset.

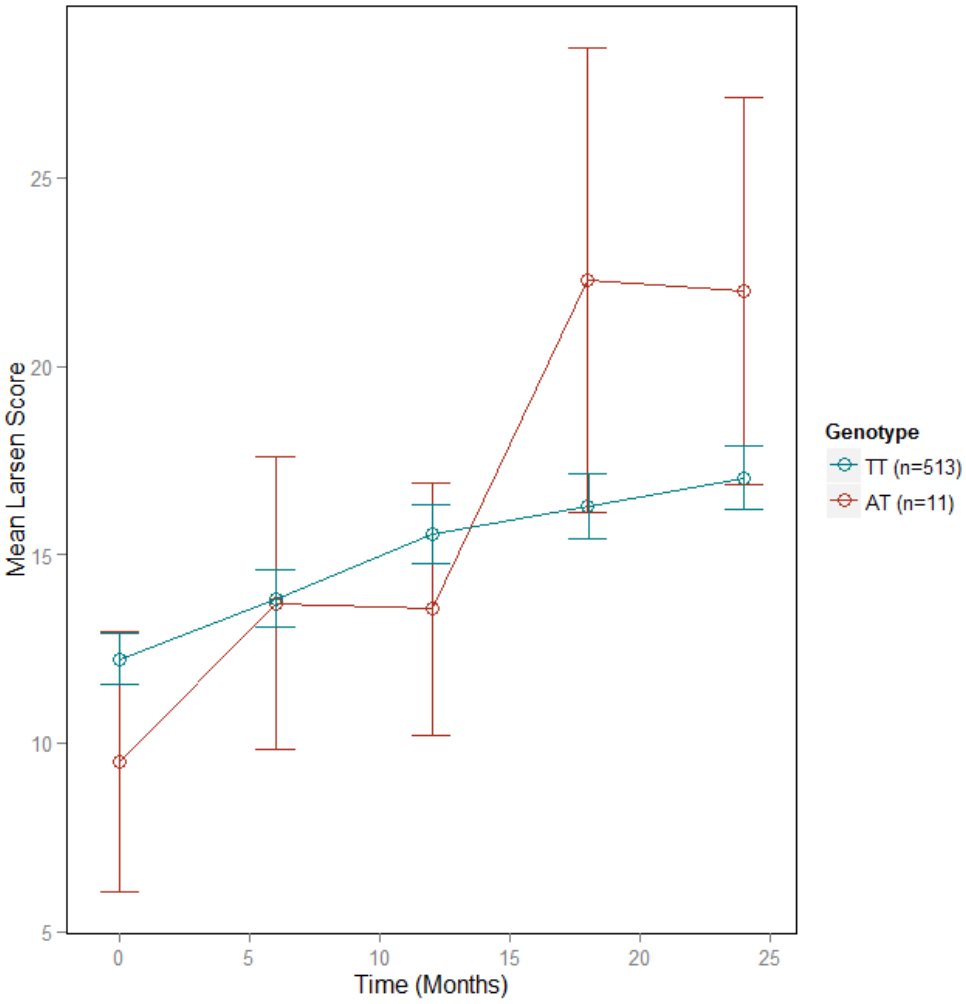
Figure 4-4. Regional Association Plot of SNPs within 400kb of rs35309890 with X-Ray Progression



*P-values for SNPs located within 400kb of rs35309890 (purple circle) on chromosome 3; LD relative to rs35309890; co-ordinates based on GRCh37 assembly; P-values from genotype*time interaction term; plot generated using Locus Zoom version 1.1 (Pruim, et al., 2010).*

The linear mixed-effects model estimated that the presence of the rs35309890 minor A allele was associated with a 1.54-fold (95% CI 1.28-1.87) greater annual increase in modified Larsen scores relative to the common T allele (Table 4-1). This association is demonstrated in Figure 4-5. In the 11 heterozygous (AT genotype) patients, the mean Larsen score increase over 24-months was 12.5 units; in the 513 homozygous (TT genotype) patients the mean Larsen score increase was 4.8 units. This represents a 1.3-fold greater annual increase in modified Larsen scores in AT heterozygous patients compared with TT homozygous patients. Due to the rarity of this SNP (MAF=0.01) the Larsen scores in AT heterozygous patients were known with substantially less precision, as reflected by the wide standard error bars in Figure 4-5. There were no AA homozygous patients.

Figure 4-5. Mean Larsen Scores Stratified By rs35309890 Genotype

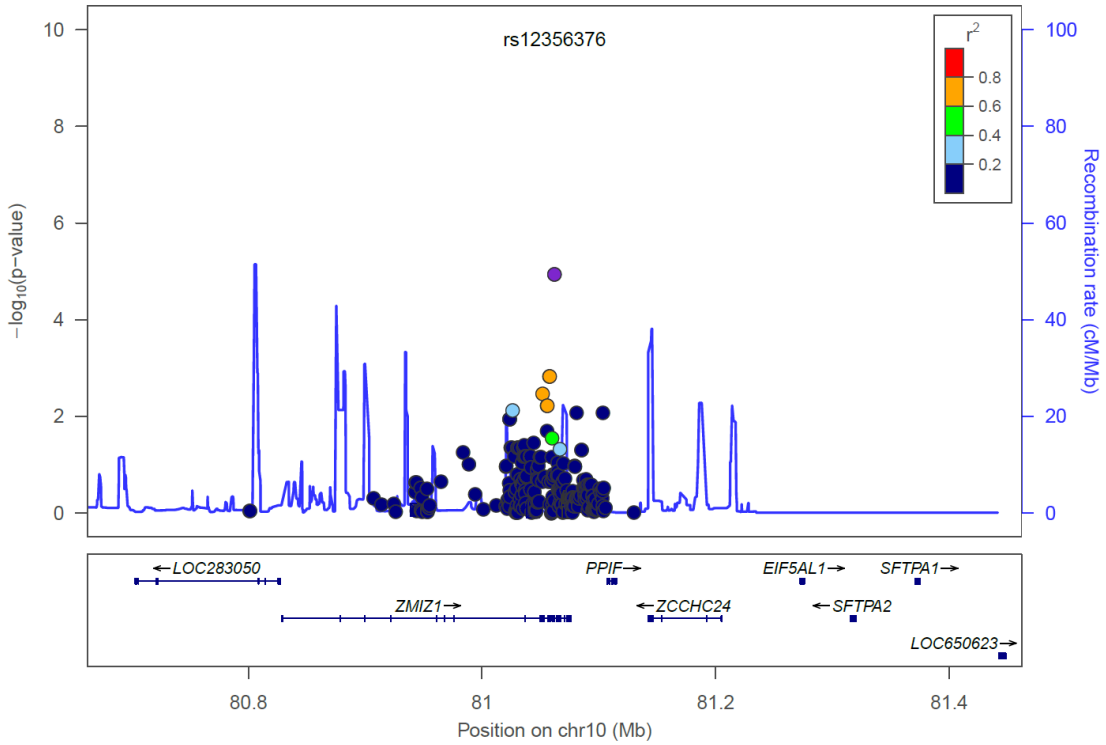


Mean Larsen scores with standard error bars at each time point; the slight decrease in mean scores in AT homozygous patients at 12 and 24 months are due to missing X-ray data in two CARDERA-2 patients at the 6 and 18 month time points.

4.3.3.2.rs12356376

The second most significant SNP was rs12356376, which resides on chromosome 10 in the *ZMIZ1* (zinc finger, MIZ-type containing 1) locus. The *ZMIZ1* gene encodes the protein zinc finger MIZ type 1, which belongs to the protein inhibitor of activated STAT (PIAS) family (Ellinghaus, et al., 2012). This protein regulates the activity of a range of transcription factors, including Smad3/4 and p53 (Ellinghaus et al, 2012). A regional plot of the association between other SNPs in this area and radiological progression (Figure 4-6) indicated that no SNPs in this region were in high LD with rs12356376.

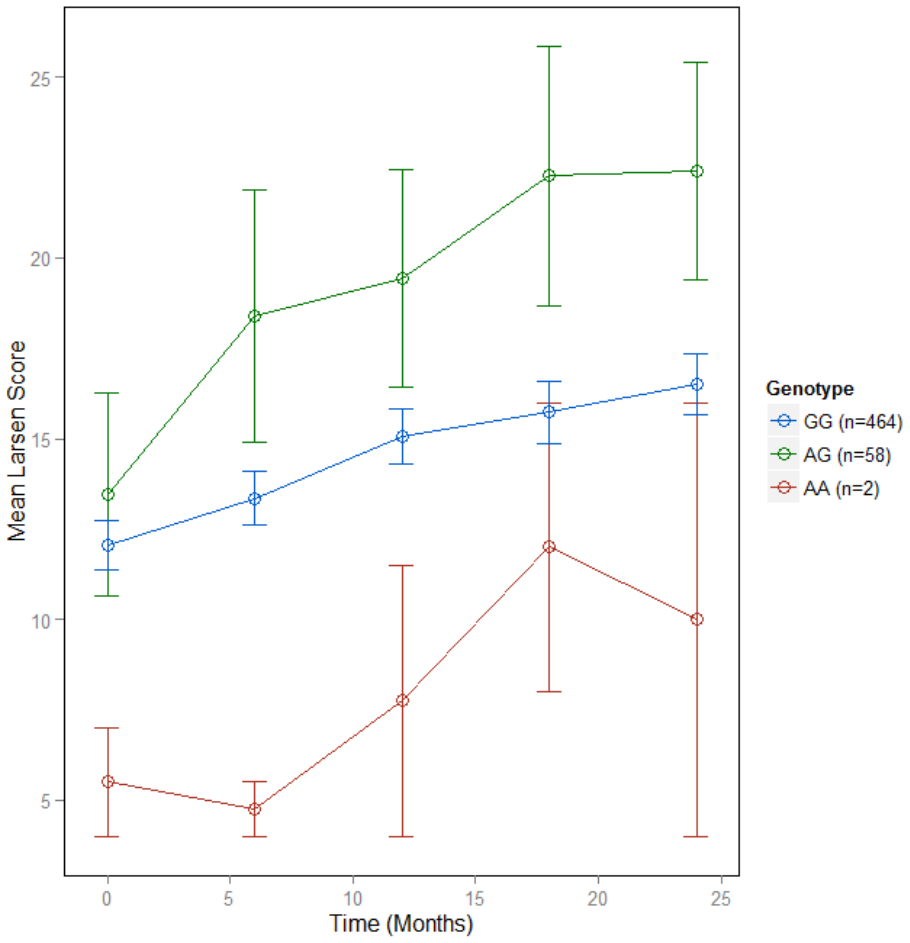
Figure 4-6. Regional Association Plot of SNPs within 400kb of rs12356376 with X-Ray Progression



*P-values for SNPs located within 400kb of rs12356376 (purple circle) on chromosome 10; LD relative to rs12356376; co-ordinates based on GRCh37 assembly; P-values from genotype*time interaction; plot generated using Locus Zoom version 1.1 (Pruim et al, 2010).*

The linear mixed-effects model estimated that the presence of the rs12356376 minor A allele associated with a 1.19-fold (95% CI 1.10-1.29) greater annual increase in modified Larsen scores relative to the common G allele (Table 4-1). Examining mean Larsen scores stratified by rs12356376 genotype (Figure 4-7) indicated this was an accurate approximation for the effect of carrying one copy of the A allele. The mean Larsen score increase over 24-months was 4.44 units in the 464 homozygous (GG genotype) patients and 8.95 in the 58 heterozygous (AG genotype) patients. This represented a 1.01-fold greater annual increase in modified Larsen scores in AG heterozygous patients compared with GG homozygous patients. No significant difference between 2-year mean changes in Larsen scores in AA versus GG homozygous patients was seen, although the relevance of this is uncertain as only 2 patients were AA homozygotes.

Figure 4-7 Mean Larsen Scores Stratified By rs12356376 Genotype



Mean Larsen scores with standard error bars shown for each time point

4.4. DISCUSSION

In this genome-wide study, which evaluated the relationship between genetic markers known to associate with immune-mediated diseases and radiological progression in early, active RA patients, no variant attained a significant association that passed the pre-defined *P*-value threshold. This negative analysis probably reflects the limited size of the study, which was only powered to detect variants of a large effect size (defined as ≥7% of the variance in X-ray progression). This supports the concept that, as with disease susceptibility, joint destruction rates in RA are likely to be defined by multiple loci of a small effect size.

The two SNPs that had the strongest associations with radiological progression were rs35309890 and rs12356376, in chromosomes 3 and 10, respectively. The nearest gene to rs35309890 is *EOMES*, which lies within 17.5 kb of this SNP. *EOMES* encodes the transcription factor eomesodermin, which has an important regulatory role in CD8+ T cell activity (Intlekofer et al, 2005). In murine studies it has been shown to invoke key attributes of effector CD8+ T cells such as release of the pro-inflammatory cytokine, interferon-gamma (IFN-γ); it is also required for their full effector differentiation (Pearce, et al., 2003). In humans, IFN-γ has also been shown to be controlled by eomesodermin (Atreya, et al., 2007). Rs35309890 could therefore influence articular damage through attenuating CD8+ T cell effector functions. The latter SNP, rs12356376, is an intron variant within the *ZMIZ1* locus. *ZMIZ1* has an established association with susceptibility to both crohn's disease and psoriasis (Ellinghaus et al, 2012), although its precise role in these diseases is unknown. It has been linked to the presence of malignancies including acute lymphoblastic leukaemia (due to a translocation with the protein tyrosine kinase *ABL1* locus on chromosome 9 (Soler, et al., 2008)) and breast, colonic, ovarian and cutaneous squamous cell carcinomas (Rogers, et al., 2013). The relevance of these two SNPs to radiological progression in RA is uncertain; neither attained statistical significance in the CGC and an association with X-ray progression has not been reported in other studies.

This analysis focussed on identifying genetic predictors of radiological progression. This is because X-ray progression is the only RA outcome demonstrated to have a heritable component (Knevel et al, 2012b). The relevance of radiological progression

to current clinical practice is somewhat uncertain for several reasons. Firstly, there is evidence that the natural history of RA is changing with lower levels of radiological damage observed in more contemporary RA cohorts. In an analysis of 5 clinical trials of anti-TNF in methotrexate-experienced RA patients, baseline radiographic scores had fallen by more than 50% over 10 years (Rahman, et al., 2011). Secondly, current treatment strategies of combination DMARDs and biologics significantly reduce radiological progression rates (Goekoop-Ruiterman, et al., 2008, Boers, et al., 1997a, Graudal and Jurgens, 2010). Major structural damage is therefore less commonly seen with modern treatments. Thirdly, in the context of clinical trials, radiological scores correlate poorly with clinical responses and improvements in physical function (Strand and Sharp, 2003); these are both more immediately relevant to patients than their X-ray scores. There is therefore a key research requirement to establish the heritability of other RA outcomes such as disease activity scores, disability levels and quality of life. If substantial heritability is demonstrated then identifying their genetic predictors could deliver personalised care that is more relevant to current clinical practice and patients.

The strengths and weaknesses of this study are broadly the same as those outlined for the analysis undertaken in Chapter 3 of this thesis. The limited power of this study is, however, substantially greater in view of the large number of genetic markers tested. There exists debate around the optimal strategy to correct for multiple testing for studies undertaken using the ImmunoChip. Eyre *et al* used the traditional P_{GWAS} threshold of $<5\times10^{-8}$ in their recent case-control association study of RA susceptibility variants (Eyre et al, 2012). De Rooy *et al* used the same approach as that undertaken in this study, by correcting for the number of independent tests after accounting for LD (De Rooy et al, 2013b). Their r^2 threshold of >0.5 for removing correlated genetic markers was lower than that used in this study. When applying this more conservative LD value to the CGC, the *P*-value threshold was $1.26\text{x}10^{-6}$ (39,610 uncorrelated SNPs); no genetic markers attained significance. There are three additional limitations to this study, beyond those discussed for the candidate gene approach in Chapter 3. Firstly, as with all genome-wide analyses it probably contains a substantial number of false-negative results due to the strict significance threshold used (Gibson, 2011). Secondly, the ImmunoChip only includes markers in

regions of the genome known to associate with immune diseases; it does not provide true genome-wide coverage and markers between the genomic regions present on this array may be more relevant to X-ray progression. Thirdly, since the ImmunoChip was designed in 2009 (Parkes et al, 2013), the number of validated susceptibility loci for the immune-mediated diseases on which it was based has rapidly expanded; these may not be captured on the array.

Further work is needed to establish the genetic basis of radiological progression in RA. In the single GWAS providing true genome-wide coverage that has been undertaken in this area to date, only one SNP (rs7607479) attained P_{GWAS} for an association with radiological damage (Knevel et al, 2013b). When this is considered alongside the current analysis and the other ImmunoChip study (De Rooy et al, 2013b), the genetic architecture of radiological progression is likely to comprise multiple loci of small effect sizes. As with RA susceptibility, the best approach to identify replicable genetic associations is, therefore, through meta-analyses of GWASs. In contrast to case-control association studies, such an approach will have a number of unique challenges with existing relevant genotyped cohorts using different X-ray scoring systems, having a variable number of X-rays available per individual and including patients of different disease durations and severities. The optimal strategy to undertake such analyses requires careful consideration.

In conclusion this study has demonstrated that no significant association exists between genetic markers present on the ImmunoChip and radiological progression in 524 patients with early, active RA. As the genetic basis of X-ray progression in RA probably comprises many loci of a small effect size, GWAS meta-analyses are required to optimise the power to identify relevant genetic variants.

CHAPTER 5. ACPA AS A TREATMENT BIOMARKER

This chapter is presented as a published paper and is a copy of the following journal publication:

Seegobin SD, Ma MH, Dahanayake C, Cope AP, Scott DL, Lewis CM, **Scott IC**. ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of a randomised controlled trial. Arthritis Res Ther 2014; 16: R13.

arthritis research&therapy

RESEARCH ARTICLE Open Access

ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of a randomized controlled trial

Seth D Seegobin[1], Margaret HY Ma[2], Chanaka Dahanayake[2], Andrew P Cope[2], David L Scott[3], Cathryn M Lewis[1] and Ian C Scott[1,2*]

Abstract

Introduction: UK guidelines recommend that all early active rheumatoid arthritis (RA) patients are offered combination disease-modifying antirheumatic drugs (DMARDs) and short-term corticosteroids. Anti-citrullinated protein antibody (ACPA)-positive and ACPA-negative RA may differ in their treatment responses. We used data from a randomized controlled trial - the Combination Anti-Rheumatic Drugs in Early RA (CARDERA) trial - to examine whether responses to intensive combination treatments in early RA differ by ACPA status.

Methods: The CARDERA trial randomized 467 early active RA patients to receive: (1) methotrexate, (2) methotrexate/ciclosporin, (3) methotrexate/prednisolone or (4) methotrexate/ciclosporin/prednisolone in a factorial-design. Patients were assessed every six months for two years. In this analysis we evaluated 431 patients with available ACPA status. To minimize multiple testing we used a mixed-effects repeated measures ANOVA model to test for an interaction between ACPA and treatment on mean changes from baseline for each outcome (Larsen, disease activity scores on a 28-joint count (DAS28), Health Assessment Questionnaire (HAQ), EuroQol, SF-36 physical component summary (PCS) and mental component summary (MCS) scores). When a significant interaction was present, mean changes in outcomes were compared by treatment group at each time point using t-tests stratified by ACPA status. Odds ratios (ORs) for the onset of new erosions with treatment were calculated stratified by ACPA.

Results: ACPA status influenced the need for combination treatments to reduce radiological progression. ACPA-positive patients had significant reductions in Larsen score progression with all treatments. ACPA-positive patients receiving triple therapy had the greatest benefits: two-year mean Larsen score increases comprised 3.66 (95% confidence interval (CI) 2.27 to 5.05) with triple therapy and 9.58 (95% CI 6.76 to 12.39) with monotherapy; OR for new erosions with triple therapy versus monotherapy was 0.32 (95% CI 0.14 to 0.72; $P = 0.003$). ACPA-negative patients had minimal radiological progression irrespective of treatment. Corticosteroid's impact on improving DAS28/PCS scores was confined to ACPA-positive RA.

Conclusions: ACPA status influences the need for combination DMARDs and high-dose tapering corticosteroids in early RA. In CARDERA, combination therapy was only required to prevent radiological progression in ACPA-positive patients; corticosteroids only provided significant disease activity and physical health improvements in ACPA-positive disease. This suggests ACPA is an important biomarker for guiding treatment decisions in early RA.

Trial registration: Current Controlled Trials ISRCTN32484878

* Correspondence: ian.scott@kcl.ac.uk
[1]Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, Great Maze Pond, 8th Floor Tower Wing, London SE1 9RT, UK
[2]Academic Department of Rheumatology, Centre for Molecular and Cellular Biology of Inflammation, 1st Floor, New Hunt's House, Guy's Campus, King's College London, Great Maze Pond, London SE1 1UL, UK
Full list of author information is available at the end of the article

Introduction

Rheumatoid arthritis (RA) is a heterogeneous disease spanning several subsets. One crucial subdivision is defined by the presence or absence of anti-citrullinated protein antibodies (ACPA), termed ACPA-positive and ACPA-negative RA, respectively [1]. ACPA-positive RA has a worse prognosis with higher rates of erosive damage [2]. It also has different risk factors than ACPA-negative RA with most genetic associations [3,4] and environmental risks, such as smoking [5] and alcohol abstinence [6], predominantly linked to ACPA-positive disease. These disparities suggest that RA ACPA subsets might respond differently to treatment [7].

Current RA management focuses on early intensive therapies, often using combinations of disease-modifying antirheumatic drugs (DMARDs) and glucocorticoids with rapid escalation to biologics in refractory cases. Guideline recommendations for the treatment of early RA differ across countries. UK guidelines from the National Institute for Health and Care Excellence (NICE) advocate that all individuals with active RA are offered combination DMARDs with short-term glucocorticoids [8]. American College of Rheumatology (ACR) guidelines suggest reserving combination DMARDs for patients with markers of severe disease, such as ACPA positivity [9]. The European League Against Rheumatism (EULAR) guidelines also suggest a stratified treatment approach, advocating biologics in patients with poor prognostic markers like ACPA that are failing to attain remission or low disease activity with an initial treatment strategy of synthetic DMARDs [10]. There are, however, insufficient data on prognostic factors in randomized controlled trials (RCTs) of combination DMARDs and biologics to know which approach is best.

We used data from an RCT of combination DMARDs and corticosteroids in early RA - the Combination Anti-Rheumatic Drugs in Early RA (CARDERA) trial [11] - to examine whether responses to intensive combination treatments differ by ACPA status. Our primary aim was to examine if combination DMARDs and corticosteroids had different effects on radiological progression in ACPA-positive and ACPA-negative RA. Our secondary aims were to evaluate if any differential effects also extended to disease activity, disability and quality of life (QoL).

Methods

Ethical approval

The CARDERA trial was approved by the South Thames Multicentre Research Ethics Committee (REC Reference: MREC (1) 99/04). Further ethical approval was obtained to process the archived serum for ACPA status from the East of England - Essex Research Ethics Committee (REC Reference: 11/EE/0544). Informed consent was obtained from all patients recruited to the CARDERA trial.

Subjects

The CARDERA trial recruited patients with early active RA (of less than two years duration) from 42 UK centers; its details have previously been reported [11,12]. Patients were randomized to one of four treatment arms: (1) monotherapy with methotrexate; (2) double therapy with methotrexate and ciclosporin; (3) double therapy with methotrexate and prednisolone; (4) triple therapy with methotrexate, ciclosporin and prednisolone. A factorial-design was adopted to allow the simultaneous evaluation of prednisolone and ciclosporin in a 2×2 design. Treatment groups were well matched with similar baseline characteristics [11]. Patients were assessed every 6 months (for 24 months). Missing data were imputed through last observations carried forward (undertaken in 19% of patients at 24 months). We restricted our current analysis to the 431 individuals (from 467 recruited) who had their sera archived at baseline and evaluable for ACPA.

Serological assessments

ACPA-status was evaluated using the commercial ELISA assay, the Axis-Shield DIASTAT anti-CCP2 test (Axis-Shield, Dundee, UK). All samples were processed in duplicate with a cut-off of >5 units/ml taken as positive in keeping with the manufacturer's instructions. Rheumatoid factor (RF) had been processed at recruiting center laboratories during the original trial.

Study treatments

Study treatments comprised: (1) methotrexate (starting at 7.5 mg/week and increasing by 2.5 mg every two weeks to a final dose of 15 mg/week); (2) "step-down" prednisolone (based on the trial by Boers *et al.* [13], comprising 60 mg/day in week 1, tapering to 7.5 mg/day in weeks 7 to 28 and thereafter further reduced and stopped by week 36) started with methotrexate; (3) ciclosporin (based on the trial by Pasero *et al.* [14], comprising 1.5 mg/kg daily initially, increased gradually to a target dose of 3 mg/kg daily) started three months after methotrexate. Prednisolone and ciclosporin were given as active tablets or placebos. Intra-articular glucocorticoids (40 mg methylprednisolone with lignocaine) were given (on no more than six occasions) as required. Intramuscular glucocorticoids were allowed but only three doses of 120 mg of depot methylprednisolone could be given in a year.

Outcomes

The following disease outcomes were assessed: (1) radiological damage - the onset of new erosions at 24 months and modified Larsen scores; (2) disease activity - disease activity scores on a 28-joint count (DAS28); (3) disability-Health Assessment Questionnaire (HAQ); (4) QoL- SF-36

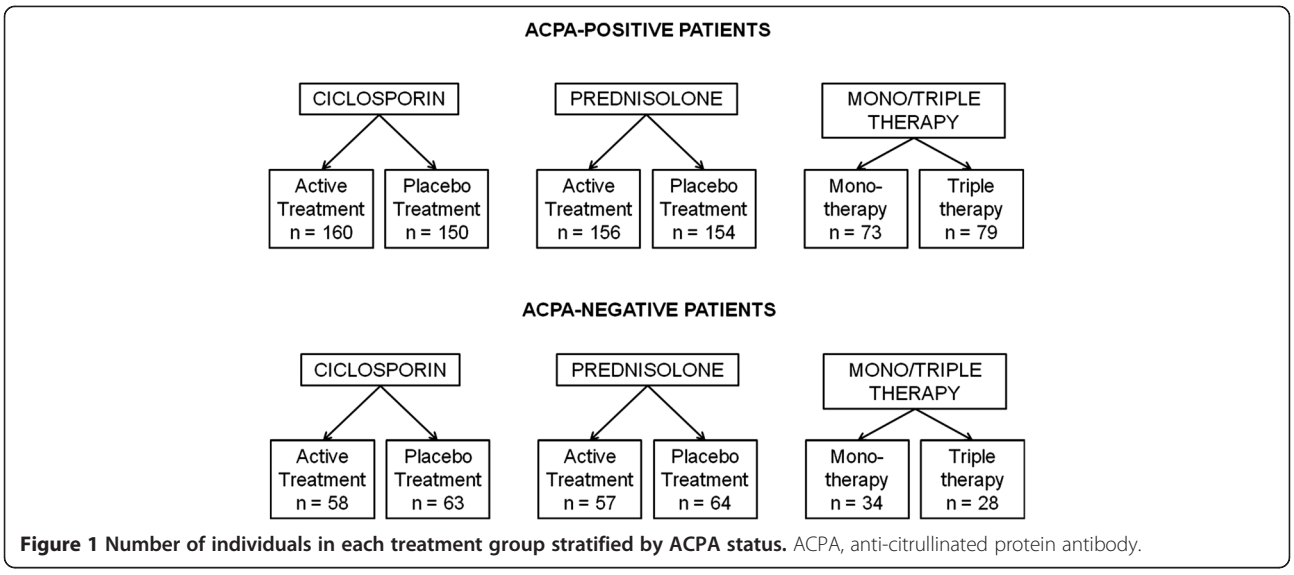


Figure 1 Number of individuals in each treatment group stratified by ACPA status. ACPA, anti-citrullinated protein antibody.

physical (PCS) and mental (MCS) component summary scores and EuroQol.

Statistical analysis

Baseline differences between ACPA-positive and ACPA-negative patients were evaluated using t-tests, Wilcoxon signed-rank tests or chi-squared tests depending on data type and distribution.

To establish whether treatment response differed by ACPA status, we used a two-staged approach; this minimized the potential for inflation of type I error associated with multiple testing. The first stage used a mixed-effects repeated measures ANOVA model to examine the effect of ACPA, treatment (coded categorically as one of the four randomized treatment arms) and time (assessment visit) on mean changes in each RA outcome (Larsen, DAS28, HAQ, EuroQol, PCS and MCS scores). The key component of this model was an ACPA*treatment interaction term, which established whether treatment responses differed by ACPA status.

The second stage was restricted to outcomes associated with significant ACPA*treatment interactions and compared mean changes in these outcomes by treatment group at each time point using t-tests in ACPA-positive and ACPA-negative patients. Where the ratio between variances significantly differed from 1, Satterthwaite's approximation was used to calculate the degrees of freedom for the critical t-statistic. This second stage allowed us to establish which treatments differed in their effects by ACPA status and how these differences changed over time. This analysis followed the original factorial grouping by comparing the following treatment groups (Figure 1): (a) active ciclosporin vs. placebo ciclosporin; (b) active prednisolone vs. placebo prednisolone; (c) triple therapy (methotrexate, cicosporin and prednisolone) vs. monotherapy (methotrexate).

Table 1 Baseline characteristics by ACPA status

Characteristic	ACPA-positive *(n = 310)*	ACPA-negative *(n = 121)*	Group difference
Female (number; %)	208 (67%)	89 (74%)	$P = 0.193^2$
RF positive (number; %)	244 (79%)	47 (39%)	$P < 0.001^2$
Age (years)	54.0 (46.0, 64.0)	55.0 (47.0, 62.0)	$P = 0.661^1$
Disease duration (months)	2.00 (0.00, 5.00)	1.00 (0.00, 4.00)	$P = 0.106^3$
Larsen score	7.50 (2.50, 21.25)*	4.50 (1.00, 9.50)	$P < 0.001^3$
DAS28	5.72 (4.91, 6.73)	5.96 (4.92, 6.85)	$P = 0.305^1$
HAQ	1.62 (1.00, 2.12)	1.62 (1.12, 2.12)	$P = 0.595^3$
EuroQol	0.60 (0.15, 0.68)	0.58 (0.08, 0.68)	$P = 0.552^3$
SF-36 PCS	28.68 (23.24, 35.95)	28.70 (22.90, 35.56)	$P = 0.713^3$
SF-36 MCS	38.64 (27.65, 53.71)	35.87 (25.37, 52.95)	$P = 0.217^3$

All data are median interquartile range (IQR) unless otherwise stated; [1] = t-test; [2] = chi-squared test; [3] = Wilcoxon signed-rank test; *baseline Larsen scores missing in two ACPA-positive patients. ACPA, anti-citrullinated protein antibody; DAS28, disease activity scores on a 28-joint count; HAQ, Health Assessment Questionnaire; MCS, mental component summary; PCS, physical component summary; RF, rheumatoid factor.

Table 2 ANOVA results for the effect of ACPA, treatment and time on changes in RA outcomes

	Larsen		DAS28		HAQ		EuroQol		PCS		MCS	
Effects	**F**	***P***	**F**	***P***	**F**	***P***	**F**	***P***	**F**	***P***	**F**	***P***
ACPA	31.90	<0.001	4.02	0.045	7.25	0.007	9.72	0.002	4.07	0.044	2.29	0.131
Time	16.83	<0.001	1.26	0.288	1.37	0.251	0.70	0.550	0.80	0.493	0.65	0.584
Treatment	9.93	<0.001	1.71	0.163	17.76	<0.001	11.47	<0.001	5.67	0.001	1.92	0.124
ACPA*Treatment	7.05	<0.001	3.99	0.008	0.48	0.696	2.94	0.032	3.22	0.022	1.84	0.138

F = F-statistic; *P* = *P*-value; ACPA*Treatment = ACPA*Treatment interaction term. ACPA, anti-citrullinated protein antibody; DAS28, disease activity scores on a 28-joint count; HAQ, Health Assessment Questionnaire; MCS, mental component summary; PCS, physical component summary; RA, rheumatoid arthritis.

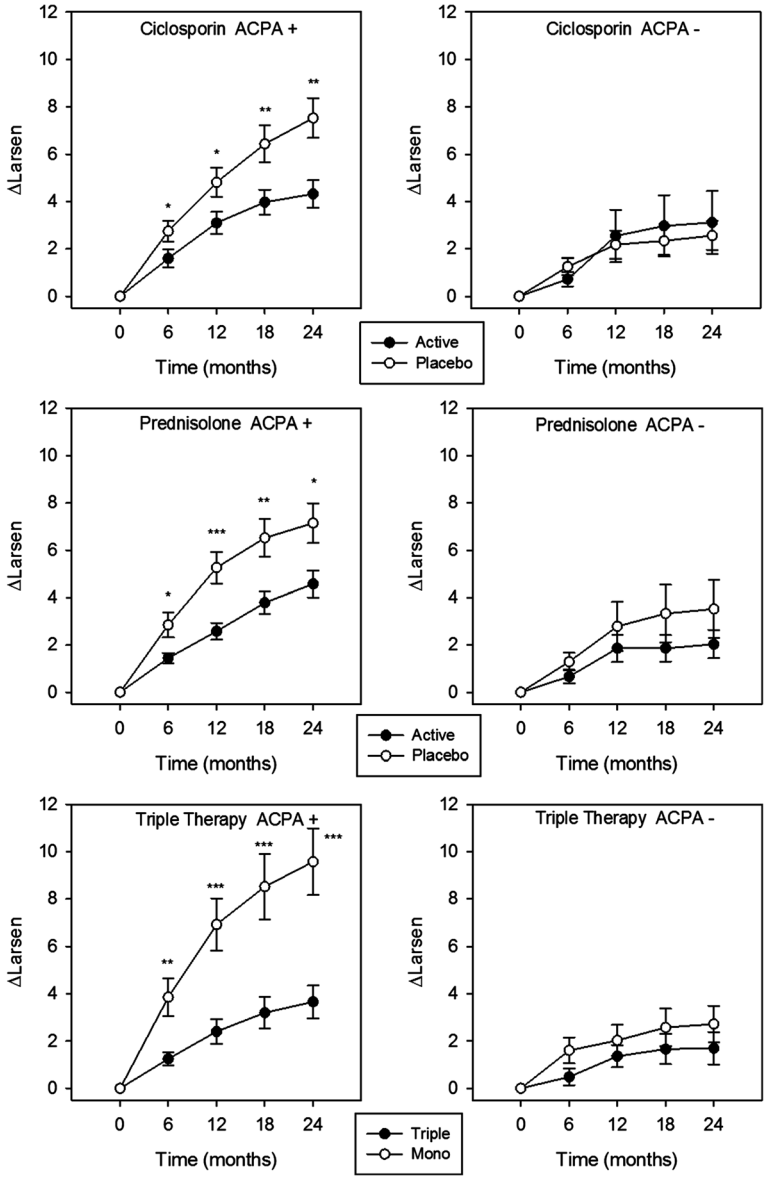


Figure 2 Treatment effect on mean changes in Larsen scores in ACPA-positive and ACPA-negative patients. Standard error bars are shown for each time point; *denotes significance at $P < 0.05$; **denotes significance at $P \leq 0.01$; ***denotes significance at $P \leq 0.001$; no asterisk denotes $P \geq 0.05$. ACPA, anti-citrullinated protein antibody.

Table 3 Treatment effects on mean changes in Larsen and DAS28 scores in ACPA-positive and ACPA-negative RA

	Time	Ciclosporin			Prednisolone			Triple vs. monotherapy		
		Ciclosporin	Placebo	*P*	Prednisolone	Placebo	*P*	Triple	Mono	*P*
Larsen										
ACPA+	6	1.60 (0.37)	2.74 (0.44)	0.047	1.44 (0.22)	2.85 (0.53)	0.015	1.25 (0.28)	3.86 (0.80)	0.003
	12	3.09 (0.47)	4.81 (0.61)	0.027	2.58 (0.35)	5.26 (0.67)	0.001	2.41 (0.52)	6.92 (1.10)	<0.001
	18	3.97 (0.53)	6.43 (0.78)	0.010	3.79 (0.48)	6.52 (0.80)	0.004	3.20 (0.66)	8.52 (1.39)	0.001
	24	4.32 (0.58)	7.53 (0.84)	0.002	4.57 (0.57)	7.15 (0.84)	0.012	3.66 (0.70)	9.58 (1.41)	<0.001
ACPA-	6	0.72 (0.30)	1.25 (0.36)	0.270	0.67 (0.29)	1.30 (0.37)	0.182	0.48 (0.35)	1.60 (0.55)	0.093
	12	2.54 (1.11)	2.17 (0.59)	0.770	1.86 (0.58)	2.79 (1.04)	0.435	1.36 (0.47)	2.03 (0.66)	0.409
	18	2.97 (1.30)	2.33 (0.60)	0.660	1.86 (0.56)	3.33 (1.21)	0.275	1.66 (0.64)	2.57 (0.79)	0.388
	24	3.11 (1.32)	2.56 (0.61)	0.704	2.04 (0.60)	3.52 (1.22)	0.277	1.70 (0.69)	2.72 (0.77)	0.335
DAS28										
ACPA+	6	−1.61 (0.12)	−1.49 (0.13)	0.488	−1.97 (0.12)	−1.13 (0.11)	<0.001	−1.98 (0.18)	−0.99 (0.17)	<0.001
	12	−1.46 (0.12)	−1.19 (0.14)	0.147	−1.36 (0.13)	−1.29 (0.13)	0.716	−1.48 (0.18)	−1.14 (0.19)	0.190
	18	−1.49 (0.13)	−1.36 (0.14)	0.498	−1.50 (0.14)	−1.36 (0.13)	0.479	−1.64 (0.20)	−1.37 (0.20)	0.356
	24	−1.62 (0.13)	−1.38 (0.15)	0.211	−1.62 (0.14)	−1.38 (0.14)	0.203	−1.84 (0.19)	−1.36 (0.22)	0.087
ACPA-	6	−1.02 (0.21)	−1.42 (0.20)	0.173	−1.47 (0.24)	−1.00 (0.17)	0.111	−1.43 (0.33)	−1.32 (0.22)	0.792
	12	−0.94 (0.21)	−1.32 (0.21)	0.209	−1.01 (0.21)	−1.25 (0.21)	0.421	−0.73 (0.33)	−1.36 (0.33)	0.186
	18	−1.33 (0.24)	−1.52 (0.20)	0.552	−1.33 (0.23)	−1.51 (0.22)	0.579	−1.27 (0.37)	−1.62 (0.30)	0.462
	24	−1.16 (0.22)	−1.49 (0.20)	0.256	−1.32 (0.20)	−1.34 (0.21)	0.960	−1.27 (0.32)	−1.59 (0.30)	0.464

Data are mean changes (SE) unless otherwise stated; *P* = *P*-values from t-tests; Triple = triple DMARD therapy; Mono = monotherapy with methotrexate; time is in months. ACPA, anti-citrullinated protein antibody; DAS28, disease activity scores on a 28-joint count.

In addition, we calculated odds ratios (ORs) for the development of new erosions with each treatment using binary logistic regression stratified by ACPA status.

P-values of <0.05 were considered significant. Analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC, USA).

Results

Patient characteristics

Of the 431 RA cases, 310 (72%) were ACPA-positive and 121 (28%) were ACPA-negative. Baseline characteristics were similar between ACPA subsets with the exception of Larsen scores and RF status (Table 1). ACPA-positive patients had more radiological damage at baseline; the difference in median Larsen scores between ACPA-subsets was 3.00 units (*P* <0.001). Significantly more ACPA-positive patients were RF-positive (*P* <0.001). Both ACPA-positive and ACPA-negative patients had median ages in the fifth decade, were mainly female, had severely active RA (median DAS28 scores >5.1) of a short duration and moderate disability (median HAQ scores 1.62). QoL was moderately impaired (median EuroQol scores 0.58 to 0.60).

Radiological progression

The first analytical step, using the ANOVA model (Table 2), showed that treatment responses differed serologically with a significant ACPA*treatment interactive effect on changes in Larsen scores observed (*P* <0.001).

The second analytical step, using the factorial approach, showed significant reductions in Larsen score progression in ACPA-positive patients receiving prednisolone, ciclosporin or triple therapy (Figure 2; Table 3). The magnitude of effect was similar with prednisolone and ciclosporin. Those receiving triple therapy had the largest reduction in radiological progression; mean Larsen score increases over 24 months were 3.66 (95% confidence interval (CI) 2.27 to 5.05) with triple therapy and 9.58 (95% CI 6.76 to 12.39) with monotherapy.

There were no significant treatment effects with any strategy in ACPA-negative patients. These individuals showed substantially less radiological progression (Figure 2; Table 3). The mean Larsen score increase in ACPA-negative patients treated with methotrexate monotherapy over 24 months was 2.72 (95% CI 1.15 to 4.29); for those receiving triple therapy the mean increase was 1.70 (95% CI 0.29 to 3.10).

Differences in radiological progression between ACPA-subsets were also seen in the proportion of patients developing new erosions (24% of ACPA-positive patients; 7% of ACPA-negative patients). Reductions in erosion development in ACPA-positive patients were similar with ciclosporin (OR 0.55; 95% CI 0.31 to 0.96; *P* = 0.032) and prednisolone (OR 0.56; 95% CI 0.32 to 0.99; *P* = 0.045) when compared with placebo. Triple therapy had the greatest

impact on reducing new erosions when compared with monotherapy (OR 0.32; 95% CI 0.14 to 0.72; $P = 0.003$). Treatment had no significant impact on preventing erosions in ACPA-negative patients. The ORs for reduction in erosion development in ACPA-negative patients comprised 0.86 (95% CI 0.16 to 4.23; $P = 1.00$) and 0.89 (95% CI 0.17 to 4.38; $P = 1.00$) with ciclosporin and prednisolone, respectively, compared to placebo and 0.79 (95% CI 0.06 to 7.53; $P = 1.00$) for triple therapy compared with monotherapy.

Disease activity

The ANOVA model (Table 2) showed a significant ACPA*-treatment interactive effect on changes in DAS28 scores ($P = 0.008$). Subsequent factorial analysis by treatment showed that prednisolone ($P < 0.001$) and triple therapy ($P < 0.001$) significantly reduced DAS28 scores at six months in ACPA-positive patients (Figure 3; Table 3). No treatment effects were seen at subsequent time points. There were no significant treatment effects in ACPA-negative patients.

Disability

The ANOVA model (Table 2) showed that although ACPA status ($P = 0.007$) and treatment ($P < 0.001$) influenced changes in HAQ scores no ACPA*treatment interaction

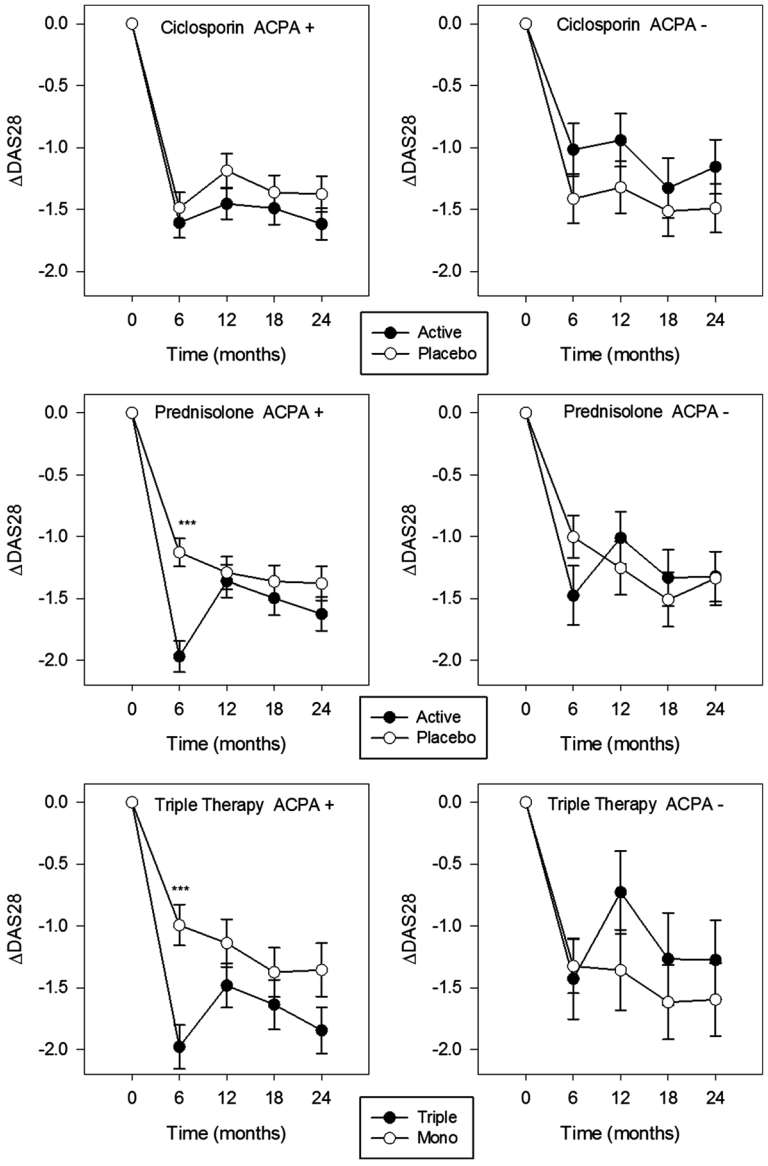


Figure 3 Treatment effect on mean changes in DAS28 scores in ACPA-positive and ACPA-negative patients. Standard error bars are shown for each time point; *denotes significance at $P < 0.05$; **denotes significance at $P \leq 0.01$; ***denotes significance at $P \leq 0.001$; no asterisk denotes $P \geq 0.05$. ACPA, anti-citrullinated protein antibody; DAS28, disease activity scores on a 28 joint count.

existed ($P = 0.696$). A factorial analysis was therefore not undertaken.

Quality of life

EuroQol

The ANOVA model (Table 2) showed a significant ACPA*-treatment interactive effect on changes in EuroQol scores ($P = 0.032$). Subsequent factorial analysis (Figure 4; Table 4) showed significant improvements in EuroQol scores at 6 months in ACPA-positive patients receiving prednisolone ($P = 0.001$) or triple therapy ($P = 0.029$); a significant effect was also seen at 12 months with ciclosporin ($P = 0.044$). A significant effect of prednisolone on EuroQol scores was seen at all time-points in ACPA-negative patients.

SF-36 PCS

The ANOVA model (Table 2) showed a significant ACPA*-treatment interactive effect on changes in PCS scores ($P = 0.022$). Factorial analysis (Figure 5; Table 4) showed that in ACPA-positive patients, prednisolone and triple therapy significantly improved PCS scores at 6 and 12 months; ciclosporin also improved PCS scores at 6 months ($P = 0.031$). In ACPA-negative patients no significant treatment effect on PCS scores was observed.

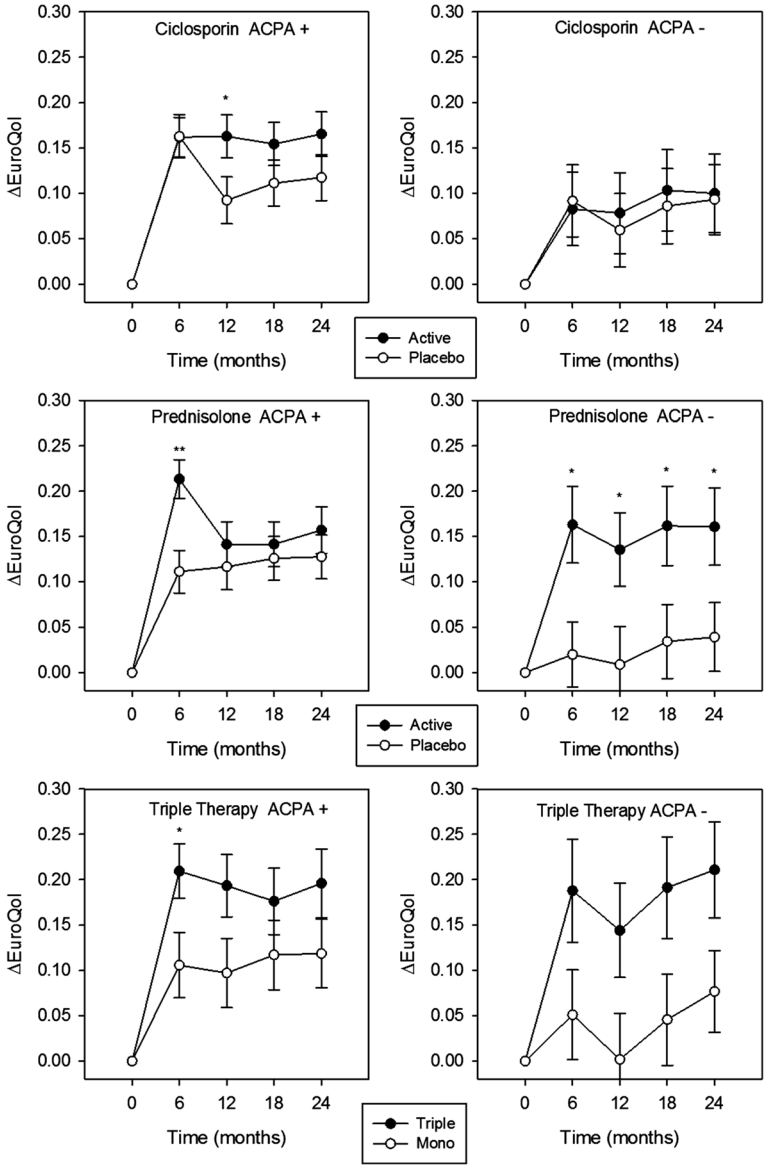


Figure 4 Treatment effect on mean changes in EuroQol scores in ACPA-positive and ACPA-negative patients. Standard error bars are shown for each time-point; *denotes significance at $P < 0.05$; **denotes significance at $P \leq 0.01$; ***denotes significance at $P \leq 0.001$; no asterisk denotes $P \geq 0.05$. ACPA, anti-citrullinated protein antibody.

Table 4 Treatment effects on mean changes in EuroQol and PCS scores in ACPA-positive and ACPA-negative RA

		Ciclosporin			Prednisolone			Triple vs. monotherapy		
	Time	Ciclosporin	Placebo	*P*	Prednisolone	Placebo	*P*	Triple	Mono	*P*
EuroQol										
ACPA+	6	0.16 (0.02)	0.16 (0.02)	0.978	0.21 (0.02)	0.11 (0.02)	0.001	0.21 (0.03)	0.11 (0.04)	0.029
	12	0.16 (0.02)	0.09 (0.03)	0.044	0.14 (0.02)	0.12 (0.02)	0.481	0.19 (0.03)	0.10 (0.04)	0.063
	18	0.15 (0.02)	0.11 (0.03)	0.215	0.14 (0.03)	0.13 (0.02)	0.651	0.18 (0.04)	0.12 (0.04)	0.267
	24	0.17 (0.02)	0.12 (0.03)	0.176	0.16 (0.03)	0.13 (0.02)	0.403	0.20 (0.04)	0.12 (0.04)	0.152
ACPA-	6	0.08 (0.04)	0.09 (0.04)	0.876	0.16 (0.04)	0.02 (0.04)	0.011	0.19 (0.06)	0.05 (0.05)	0.074
	12	0.08 (0.04)	0.06 (0.04)	0.756	0.14 (0.04)	0.01 (0.04)	0.033	0.14 (0.05)	0.00 (0.05)	0.057
	18	0.10 (0.04)	0.09 (0.04)	0.776	0.16 (0.04)	0.03 (0.04)	0.036	0.19 (0.06)	0.05 (0.05)	0.058
	24	0.10 (0.04)	0.09 (0.04)	0.908	0.16 (0.04)	0.04 (0.04)	0.035	0.21 (0.05)	0.08 (0.05)	0.057
PCS										
ACPA+	6	7.96 (0.89)	5.15 (0.95)	0.031	9.15 (0.96)	4.01 (0.83)	<0.001	10.42 (1.30)	2.29 (1.16)	<0.001
	12	5.27 (0.84)	4.17 (0.94)	0.380	6.04 (0.93)	3.42 (0.84)	0.037	7.05 (1.23)	3.28 (1.27)	0.035
	18	6.11 (0.89)	3.87 (1.02)	0.097	5.51 (1.01)	4.53 (0.89)	0.472	7.77 (1.39)	4.58 (1.44)	0.114
	24	5.77 (0.96)	4.91 (1.03)	0.544	6.36 (1.03)	4.34 (0.95)	0.150	8.33 (1.46)	5.53 (1.49)	0.180
ACPA-	6	2.64 (1.41)	4.04 (1.21)	0.449	4.14 (1.31)	2.68 (1.29)	0.428	4.86 (2.11)	4.54 (1.78)	0.908
	12	3.27 (1.35)	4.93 (1.15)	0.349	3.70 (1.25)	4.52 (1.25)	0.646	3.25 (2.07)	5.61 (1.76)	0.386
	18	3.86 (1.38)	5.06 (1.24)	0.516	4.41 (1.40)	4.55 (1.23)	0.937	4.72 (2.11)	5.88 (1.67)	0.664
	24	4.65 (1.69)	5.17 (1.07)	0.793	5.51 (1.58)	4.40 (1.21)	0.575	7.44 (2.70)	6.48 (1.36)	0.753

Data are mean changes (SE) unless otherwise stated; *P* = *P*-values from t-tests; Triple = triple DMARD therapy; Mono = monotherapy with methotrexate; time is in months. ACPA, anti-citrullinated protein antibody; PCS, physical component summary.

SF-36 MCS

The ANOVA model (Table 2) showed no significant associations between ACPA, time or treatment and MCS scores; no ACPA*treatment interaction was observed (*P* = 0.138). Factorial analysis was therefore not undertaken.

Discussion

Our main finding is that combination DMARDs and high-dose tapering corticosteroids are only required to prevent radiological progression in patients with early active RA in whom ACPA is present. In ACPA-positive patients, methotrexate monotherapy resulted in considerable worsening of radiological damage; the average annual Larsen score increase was 4.8 units and 38% developed new erosions. This was significantly reduced with combination treatment; in ACPA-positive patients receiving triple therapy the average annual Larsen score increase was 1.8 units and 16% developed new erosions. In contrast, ACPA-negative patients had minimal radiological progression irrespective of the treatment strategy used; the average annual increases in Larsen scores were below the minimal clinically important difference (MCID) of 2.3 units [15] with all treatments and only 7% developed new erosions.

Our other finding was that the beneficial effect of high-dose corticosteroids on reducing disease activity and improving physical health was also confined to ACPA-positive RA. Only ACPA-positive patients had significant six-month improvements in DAS28 and PCS scores with double and triple therapy regimens incorporating prednisolone. Our findings are consistent with the IMPROVED study, which also found that high-dose corticosteroids had a significantly larger effect on improving disease activity and remission rates in ACPA-positive, as compared to ACPA-negative, inflammatory arthritis patients [16]. The mechanism underlying this differential steroid response is uncertain. The fact these improvements were not maintained over time in CARDERA is expected and consistent with the original COBRA study [13]. Our results support the use of high-dose tapering corticosteroids as a bridging therapy in early RA but suggest this treatment strategy would be best reserved for ACPA-positive patients.

The impact of ACPA status on EuroQol responses to combination DMARDs and corticosteroids was less clear, with similar EuroQol improvements observed in ACPA-positive and ACPA-negative patients receiving active prednisolone and triple therapy. Interestingly, ACPA-negative patients receiving placebo prednisolone or methotrexate monotherapy had substantially smaller EuroQol improvements (maximal increase of 0.04 and 0.08 units, respectively) when compared to ACPA-positive patients (maximal increase of 0.13 and 0.12 units, respectively). This suggests that methotrexate monotherapy could be more effective at improving QoL in ACPA-positive disease.

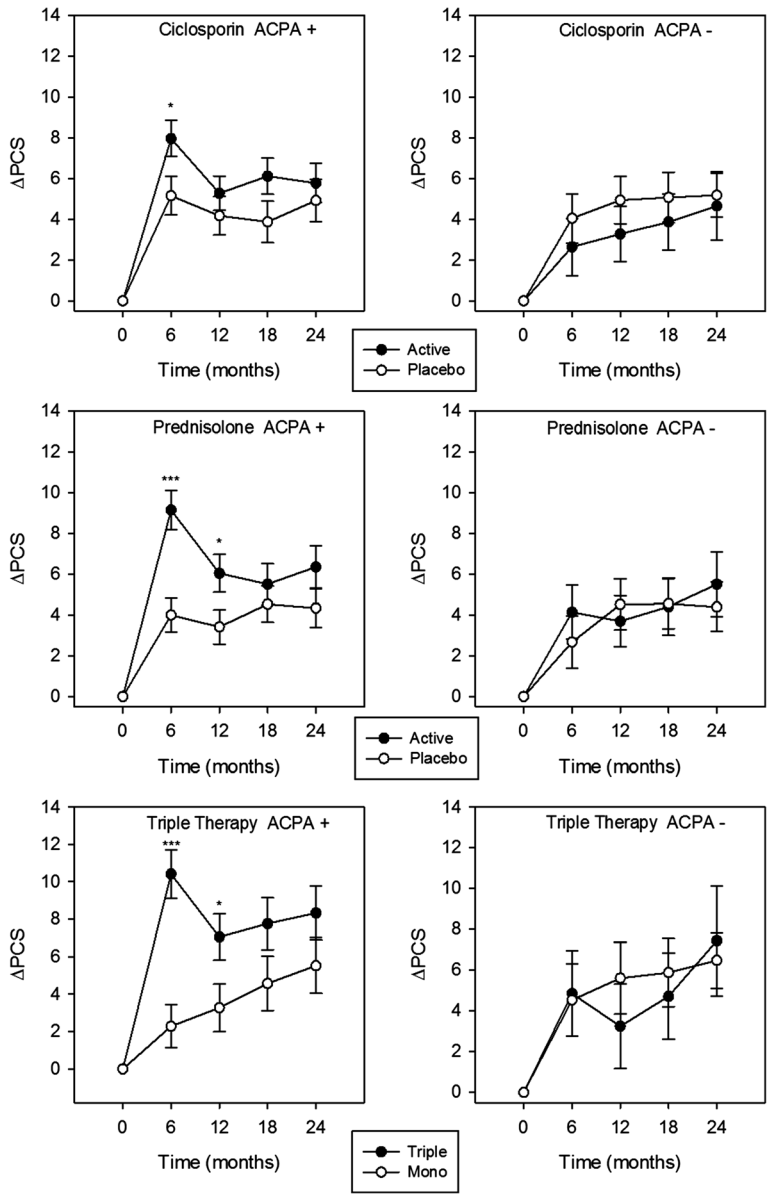


Figure 5 Treatment effect on mean changes in PCS scores in ACPA-positive and ACPA-negative patients. Standard error bars are shown for each time point; *denotes significance at $P < 0.05$; **denotes significance at $P \leq 0.01$; ***denotes significance at $P \leq 0.001$; no asterisk denotes $P \geq 0.05$. ACPA, anti-citrullinated protein antibody; PCS, physical component summary.

To our knowledge, research on the impact of ACPA status on responses to combination DMARDs and corticosteroids is limited to secondary analyses or extension studies of the BeST [17,18] and FIN-RACo [19] trials. Data from the BeST study support our finding that DMARD monotherapy is inadequate at preventing radiological progression in ACPA-positive RA; the presence of ACPA associated with radiological progression in individuals receiving monotherapy (OR for radiological progression: 12.6; 95% CI 3.0 to 51.9) but not combination therapy with DMARDs and corticosteroids (OR 1.7; 95% CI 0.5 to 5.4) [17]. This study also reported minimal radiological progression in ACPA-negative patients in all treatment groups. Although the FIN-RACo trial found that combination therapy reduced radiological progression in ACPA-negative, but not ACPA-positive patients, the study had a small sample size, allowed corticosteroids in both treatment arms and had no treatment restrictions for the last three years of follow-up [19]. The impact of ACPA-status on biologic responses has been studied in greater detail, particularly in national registries. Anti-tumor necrosis factor (anti-TNF) therapies appear more effective in ACPA-negative disease [20,21]. Conversely, T-cell and B-cell inhibition with abatacept and rituximab, respectively, appear

more efficacious in ACPA-positive RA [22,23]. Taken together these findings suggest that treatment responses differ between ACPA-positive and ACPA-negative RA. This highlights a requirement for future RCTs of RA treatments to stratify their analyses by ACPA status.

Our results show that ACPA is an important prognostic biomarker in early RA, with its presence signaling a requirement for intensive combination treatment. The heterogeneous nature of RA alongside the increasing breadth of available therapies means that identifying predictors of treatment responses is a key research goal. Studies have identified several clinical parameters associated with good anti-TNF outcomes; these include not smoking, male gender and a younger age [20,24]. Genetic markers also offer promise with a recent large genome-wide association study reporting the first definitive genetic association (in the *CD84* gene) with anti-TNF response [25]. Other smaller studies suggest that stimulated whole blood cell pro-inflammatory cytokine levels [26] and serum proteins [27] may be useful in predicting anti-TNF efficacy. These findings are promising but lack clinical utility, since most markers require validation in larger cohorts or associate with only small differences in treatment response. Further work is required to identify predictors of treatment responses in RA.

Our study has a number of strengths. These include its large sample size, the involvement of multiple centers, the measurement of a wide range of outcomes and the use of two-year follow-up data. It also has several limitations. It was a secondary analysis of a published RCT and, therefore, neither its primary hypothesis nor its statistical analysis plan was pre-specified. ACPA status was unknown in 8% of patients, who were excluded from our analysis. One DMARD, ciclosporin, is not widely used in current practice. Fewer ACPA-negative patients were studied; however, the power to detect a MCID in Larsen scores between combination therapy and monotherapy treatment arms in ACPA-negative patients was higher (86%) than in ACPA-positive patients (55%). Finally, the maximal dose of methotrexate was 15 mg/week; higher doses are often used in contemporary clinical care [28].

Different guidelines, constructed using the same evidence base, have drawn alternative conclusions regarding the optimal treatment of early active RA. NICE guidelines recommend offering all patients combination DMARDs and short-term corticosteroids [8]. ACR and EULAR guidelines recommend adopting an individualized approach to treatment intensity based on prognostic factors, such as ACPA [9,10]. Our findings favor this latter approach. They show strong evidence that ACPA-positive patients benefit from intensive combination therapy but no evidence that combination treatments improve disease outcomes beyond methotrexate monotherapy in ACPA-negative patients. We recommend that future trials in early RA should consider ACPA status when evaluating treatment outcomes. When NICE and other clinical guidelines are updated, the heterogeneity of RA requires consideration, particularly the impact of ACPA-status on treatment requirements and responses.

Conclusions

We have demonstrated that the requirement for, and response to, combination DMARDs and high-dose tapering corticosteroids differs between patients with ACPA-positive and ACPA-negative early RA. In our study, intensive combination therapy was only needed to prevent radiological progression in ACPA-positive patients. Additionally, corticosteroids only provided significant improvements in disease activity and physical health outcomes in ACPA-positive RA. These findings suggest that ACPA is an important biomarker for guiding treatment decisions in early RA. They support ACR and EULAR RA management guidelines, which recommend an individualized approach to treatment intensity based on prognostic factors such as ACPA.

Abbreviations

ACPA: Anti-citrullinated protein antibodies; ACR: American College of Rheumatology; anti-TNF: Anti-tumor necrosis factor; CARDERA: Combination Anti-Rheumatic Drugs in Early RA; CI: Confidence interval; DAS28: Disease activity score on a 28-joint count; DMARD: Disease-modifying anti-rheumatic drug; EULAR: The European League Against Rheumatism; HAQ: Health assessment questionnaire; MCID: Minimal clinically important difference; MCS: Mental component summary; NICE: National Institute for Health and Care Excellence; OR: Odds ratio; PCS: Physical component summary; QoL: Quality of life; RA: Rheumatoid arthritis; RCT: Randomised controlled trial; RF: Rheumatoid factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ICS and DLS conceived and designed the study. ICS, MHYM and CD carried out the ACPA ELISA assays. SDS, ICS and CML performed the statistical analysis. SDS, ICS, CML and APC interpreted the data. ICS, SDS, CML, APC and DLS drafted the manuscript. All authors revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by Arthritis Research UK (Grant Reference Number 19739 to ICS). It represents research arising from an Arthritis Research UK funded Clinical Research Fellowship (ICS). It also presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-1208-18256). MHYM is a recipient of an NIHR Doctoral Research Fellowship (Grant Reference Number DRF-2009-02-86). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. We also acknowledge support from the NIHR Biomedical Research Centre at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. The funders had no role in the study design, data collection and analysis, data interpretation, the writing of the manuscript or the decision to submit the manuscript for publication.

Author details

[1]Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, Great Maze Pond, 8th Floor Tower Wing, London SE1 9RT, UK. [2]Academic Department of Rheumatology, Centre for Molecular and Cellular

Biology of Inflammation, 1st Floor, New Hunt's House, Guy's Campus, King's College London, Great Maze Pond, London SE1 1UL, UK. [3]Department of Rheumatology, 3rd Floor, Weston Education Centre, King's College Hospital, Cutcombe Road, London SE5 9RJ, UK.

Received: 15 September 2013 Accepted: 27 December 2013
Published: 16 January 2014

References

1. Scott DL, Wolfe F, Huizinga TW: **Rheumatoid arthritis.** *Lancet* 2010, **376:**1094–1108.
2. van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Toes RE, Huizinga TW: **Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis.** *Arthritis Res Ther* 2005, **7:**R949–R958.
3. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, Zhernakova A, Stahl E, Viatte S, McAllister K, Amos CI, Padyukov L, Toes RE, Huizinga TW, Wijmenga C, Trynka G, Franke L, Westra HJ, Alfredsson L, Hu X, Sandor C, de Bakker PI, Davila S, Khor CC, Heng KK, Andrews R, Edkins S, Hunt SE, Langford C, Symmons D, *et al*: **High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis.** *Nat Genet* 2012, **44:**1336–1340.
4. Padyukov L, Seielstad M, Ong RT, Ding B, Ronnelid J, Seddighzadeh M, Alfredsson L, Klareskog L: **Epidemiological Investigation of Rheumatoid Arthritis study group: a genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis.** *Ann Rheum Dis* 2011, **70:**259–265.
5. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, Ronnelid J, Harris HE, Ulfgren A-K, Rantapaa-Dahlqvist S, Eklund A, Padyukov L, Alfredsson L: **A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination.** *Arthritis Rheum* 2006, **54:**38–46.
6. Scott IC, Tan R, Stahl D, Steer S, Lewis CM, Cope AP: **The protective effect of alcohol on developing rheumatoid arthritis: a systematic review and meta-analysis.** *Rheumatology* 2013, **52:**856–867.
7. Daha NA, Toes RE: **Rheumatoid arthritis: are ACPA-positive and ACPA-negative RA the same disease?** *Nat Rev Rheumatol* 2011, **7:**202–203.
8. Deighton C, O'Mahony R, Tosh J, Turner C, Rudolf M: **Guideline Development Group: management of rheumatoid arthritis: summary of NICE guidance.** *BMJ* 2009, **338:**b702.
9. Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM, Moreland LW, O'Dell J, Winthrop KL, Beukelman T, Bridges SL Jr, Chatham WW, Paulus HE, Suarez-Almazor M, Bombardier C, Dougados M, Khanna D, King CM, Leong AL, Matteson EL, Schousboe JT, Moynihan E, Kolba KS, Jain A, Volkmann ER, Agrawal H, Bae S, Mudano AS, Patkar NM, Saag KG: **2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis.** *Arthritis Care Res* 2012, **64:**625–639.
10. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, Gorter S, Knevel R, Nam J, Schoels M, Aletaha D, Buch M, Gossec L, Huizinga T, Bijlsma JW, Burmester G, Combe B, Cutolo M, Gabay C, Gomez-Reino J, Kouloumas M, Kvien TK, Martin-Mola E, McInnes I, Pavelka K, van Riel P, Scholte M, Scott DL, Sokka T, Valesini G, *et al*: **EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs.** *Ann Rheum Dis* 2010, **69:**964–975.
11. Choy EH, Smith CM, Farewell V, Walker D, Hassell A, Chau L, Scott DL, CARDERA (Combination Anti-Rheumatic Drugs in Early Rheumatoid Arhritis) Trial Group: **Factorial randomised controlled trial of glucocorticoids and combination disease modifying drugs in early rheumatoid arthritis.** *Ann Rheum Dis* 2008, **67:**656–663.
12. Ma MH, Ibrahim F, Walker D, Hassell A, Choy EH, Kiely PD, Williams R, Walsh DA, Young A, Scott DL: **Remission in early rheumatoid arthritis: predicting treatment response.** *J Rheumatol* 2012, **39:**470–475.
13. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, van Denderen JC, van Zeben D, Dijkmans BA, Peeters AJ, Jacobs P, van den Brink HR, Schouten HJ, van der Heijde DM, Boonen A, van der Linden S: **Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis.** *Lancet* 1997, **350:**309–318.
14. Pasero G, Priolo F, Marubini E, Fantini F, Ferraccioli G, Magaro M, Marcolongo R, Oriente P, Pipitone V, Portioli I, Tirri G, Trotta F, Della Casa-Alberighi O: **Slow progression of joint damage in early rheumatoid arthritis treated with cyclosporin A.** *Arthritis Rheum* 1996, **39:**1006–1015.
15. Bruynesteyn K, van der Heijde D, Boers M, Saudan A, Peloso P, Paulus H, Houben H, Griffiths B, Edmonds J, Bresnihan B, Boonen A, van der Linden S: **Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference.** *Arthritis Rheum* 2002, **46:**913–920.
16. Wevers-de Boer K, Visser K, Heimans L, Ronday HK, Molenaar E, Groenendael JH, Peeters AJ, Westedt M-L, Collée G, de Sonnaville PBJ, Grillet BA, Huizinga TW, Allaart CF: **Remission induction therapy with methotrexate and prednisone in patients with early rheumatoid and undifferentiated arthritis (the IMPROVED study).** *Ann Rheum Dis* 2012, **71:**1472–1477.
17. de Vries-Bouwstra JK, Goekoop-Ruiterman YP, Verpoort KN, Schreuder GM, Ewals JA, Terwiel JP, Ronday HK, Kerstens PJ, Toes RE, de Vries RR, Breedveld FC, Dijkmans BA, Huizinga TW, Allaart CF: **Progression of joint damage in early rheumatoid arthritis: association with HLA-DRB1, rheumatoid factor, and anti-citrullinated protein antibodies in relation to different treatment strategies.** *Arthritis Rheum* 2008, **58:**1293–1298.
18. van den Broek M, Dirven L, Klarenbeek NB, Molenaar TH, Han KH, Kerstens PJ, Huizinga TW, Dijkmans BA, Allaart CF: **The association of treatment response and joint damage with ACPA-status in recent-onset RA: a subanalysis of the 8-year follow-up of the BeSt study.** *Ann Rheum Dis* 2012, **71:**245–248.
19. Mustila A, Korpela M, Haapala AM, Kautiainen H, Laasonen L, Mottonen T, Leirisalo-Repo M, Ilonen J, Jarvenpaa S, Luukkainen R, Hannonen P: **Anti-citrullinated peptide antibodies and the progression of radiographic joint erosions in patients with early rheumatoid arthritis treated with FIN-RACo combination and single disease-modifying antirheumatic drug strategies.** *Clin Exp Rheumatol* 2011, **29:**500–505.
20. Canhao H, Rodrigues AM, Mourao AF, Martins F, Santos MJ, Canas-Silva J, Polido-Pereira J, Pereira Silva JA, Costa JA, Araujo D, Silva C, Santos H, Duarte C, da Silva JA, Pimentel-Santos FM, Branco JC, Karlson EW, Fonseca JE, Solomon DH: **Comparative effectiveness and predictors of response to tumour necrosis factor inhibitor therapies in rheumatoid arthritis.** *Rheumatology* 2012, **51:**2020–2026.
21. Potter C, Hyrich KL, Tracey A, Lunt M, Plant D, Symmons DP, Thomson W, Worthington J, Emery P, Morgan AW, Wilson AG, Isaacs J, Barton A, BRAGGSS: **Association of rheumatoid factor and anti-cyclic citrullinated peptide positivity, but not carriage of shared epitope or PTPN22 susceptibility variants, with anti-tumour necrosis factor response in rheumatoid arthritis.** *Ann Rheum Dis* 2009, **68:**69–74. Erratum in: *Ann Rheum Dis* 2011, **70:**1519.
22. Gottenberg JE, Ravaud P, Cantagrel A, Combe B, Flipo RM, Schaeverbeke T, Houvenagel E, Gaudin P, Loeuille D, Rist S, Dougados M, Sibilia J, Le Loet X, Marcelli C, Bardin T, Pane I, Baron G, Mariette X: **Positivity for anti-cyclic citrullinated peptide is associated with a better response to abatacept: data from the 'Orencia and Rheumatoid Arthritis'registry.** *Ann Rheum Dis* 2012, **71:**1815–1819.
23. Isaacs JD, Cohen SB, Emery P, Tak PP, Wang J, Lei G, Williams S, Lal P, Read SJ: **Effect of baseline rheumatoid factor and anticitrullinated peptide antibody serotype on rituximab clinical response: a meta-analysis.** *Ann Rheum Dis* 2013, **72:**329–336.
24. Kleinert S, Tony H-P, Krause A, Feuchtenberger M, Wassenberg S, Richter C, Rother E, Spieler W, Gnann H, Wittig BM: **Impact of patient and disease characteristics on therapeutic success during adalimumab treatment of patients with rheumatoid arthritis: data from a German noninterventional observational study.** *Rheumatol Int* 2012, **32:**2759–2767.
25. Cui J, Stahl EA, Saevarsdottir S, Miceli C, Diogo D, Trynka G, Raj T, Mirkov MU, Canhao H, Ikari K, Terao C, Okada Y, Wedren S, Askling J, Yamanaka H, Momohara S, Taniguchi A, Ohmura K, Matsuda F, Mimori T, Gupta N, Kuchroo M, Morgan AW, Isaacs JD, Wilson AG, Hyrich KL, Herenius M, Doorenspleet ME, Tak P-P, Crusius JB, *et al*: **Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis.** *PLoS Genet* 2013, **9:**e1003394.
26. Kayakabe K, Kuroiwa T, Sakurai N, Ikeuchi H, Kadiombo AT, Sakairi T, Kaneko Y, Maeshima A, Hiromura K, Nojima Y: **Interleukin-1 measurement in stimulated whole blood cultures is useful to predict response to anti-TNF therapies in rheumatoid arthritis.** *Rheumatology* 2012, **51:**1639–1643.
27. Ortea I, Roschitzki B, Ovalles JG, Longo JL, de la Torre I, Gonzalez I, Gomez-Reino JJ, Gonzalez A: **Discovery of serum proteomic biomarkers**

for prediction of response to infliximab (a monoclonal anti-TNF antibody) treatment in rheumatoid arthritis: an exploratory analysis. *J Proteomics* 2012, **77**:372–382.

28. Visser K, van der Heijde D: **Optimal dosage and route of administration of methotrexate in rheumatoid arthritis: a systematic review of the literature.** *Ann Rheum Dis* 2009, **68**:1094–1099.

doi:10.1186/ar4439
Cite this article as: Seegobin *et al.*: ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of a randomized controlled trial. *Arthritis Research & Therapy* 2014 **16**:R13.

CHAPTER 6. PREDICTING RHEUMATOID ARTHRITIS

This chapter is presented as a published paper and is a copy of the following journal publication:

Scott IC, Seegobin SD, Steer S, Tan R, Forabosco P, Hinks A, Eyre S, Morgan AW, Wilson AG, Hocking LJ, Wordsworth P, Barton A, Worthington J, Cope AP, Lewis CM. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. PLoS Genet 2013; 9: e1003808.

Predicting the Risk of Rheumatoid Arthritis and Its Age of Onset through Modelling Genetic Risk Variants with Smoking

Ian C. Scott[1,2]*, Seth D. Seegobin[2], Sophia Steer[3], Rachael Tan[1], Paola Forabosco[4], Anne Hinks[5], Stephen Eyre[5], Ann W. Morgan[6], Anthony G. Wilson[7], Lynne J. Hocking[8], Paul Wordsworth[9], Anne Barton[5], Jane Worthington[5], Andrew P. Cope[1], Cathryn M. Lewis[2,10]

1 Academic Department of Rheumatology, Centre for Molecular and Cellular Biology of Inflammation, King's College London, London, United Kingdom, **2** Department of Medical and Molecular Genetics, King's College London, London, United Kingdom, **3** Department of Rheumatology, King's College Hospital, London, United Kingdom, **4** Istituto di Genetica delle Popolazioni, Consiglio Nazionale delle Ricerche, Sassari, Italy, **5** Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, The University of Manchester, Manchester, United Kingdom, **6** Division of Musculoskeletal Disease, Leeds Institute of Molecular Medicine, University of Leeds and National Institute for Health Research – Leeds Musculoskeletal Biomedical Research Unit, Leeds, United Kingdom, **7** Academic Unit of Rheumatology, Department of Infection and Immunity, University of Sheffield Medical School, Sheffield, United Kingdom, **8** Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, United Kingdom, **9** NIHR Oxford Musculoskeletal BRU, Nuffield Orthopaedic Centre, Oxford, United Kingdom, **10** Social, Genetic and Developmental Psychiatry Centre (MRC), Institute of Psychiatry, London, United Kingdom

Abstract

The improved characterisation of risk factors for rheumatoid arthritis (RA) suggests they could be combined to identify individuals at increased disease risks in whom preventive strategies may be evaluated. We aimed to develop an RA prediction model capable of generating clinically relevant predictive data and to determine if it better predicted younger onset RA (YORA). Our novel modelling approach combined odds ratios for 15 four-digit/10 two-digit *HLA-DRB1* alleles, 31 single nucleotide polymorphisms (SNPs) and ever-smoking status in males to determine risk using computer simulation and confidence interval based risk categorisation. Only males were evaluated in our models incorporating smoking as ever-smoking is a significant risk factor for RA in men but not women. We developed multiple models to evaluate each risk factor's impact on prediction. Each model's ability to discriminate anti-citrullinated protein antibody (ACPA)-positive RA from controls was evaluated in two cohorts: Wellcome Trust Case Control Consortium (WTCCC: 1,516 cases; 1,647 controls); UK RA Genetics Group Consortium (UKRAGG: 2,623 cases; 1,500 controls). HLA and smoking provided strongest prediction with good discrimination evidenced by an HLA-smoking model area under the curve (AUC) value of 0.813 in both WTCCC and UKRAGG. SNPs provided minimal prediction (AUC 0.660 WTCCC/0.617 UKRAGG). Whilst high individual risks were identified, with some cases having estimated lifetime risks of 86%, only a minority overall had substantially increased odds for RA. High risks from the HLA model were associated with YORA ($P<0.0001$); ever-smoking associated with older onset disease. This latter finding suggests smoking's impact on RA risk manifests later in life. Our modelling demonstrates that combining risk factors provides clinically informative RA prediction; additionally HLA and smoking status can be used to predict the risk of younger and older onset RA, respectively.

Citation: Scott IC, Seegobin SD, Steer S, Tan R, Forabosco P, et al. (2013) Predicting the Risk of Rheumatoid Arthritis and Its Age of Onset through Modelling Genetic Risk Variants with Smoking. PLoS Genet 9(9): e1003808. doi:10.1371/journal.pgen.1003808

Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received March 27, 2013; **Accepted** August 5, 2013; **Published** September 19, 2013

Funding: This manuscript was undertaken as part of an Arthritis Research UK funded Clinical Research Fellowship (ICS; Grant Reference Number 19739). Funding for the WTCCC project was provided by the Wellcome Trust (http://www.wellcome.ac.uk) under award 076113, 085475 and 090355. This work made use of data and samples generated by the 1958 Birth Cohort (NCDS) (BRIF4130). Access to these resources was enabled via the 58READIE Project funded by Wellcome Trust and Medical Research Council (grant numbers WT095219MA and G1001799). A full list of the financial, institutional and personal contributions to the development of the 1958 Birth Cohort Biomedical resource is available at http://www2.le.ac.uk/projects/birthcohort. ICS and APC are supported by Arthritis Research UK (http://www.arthritisresearchuk.org) funding as are AH, SE, JW and AB (Grant Reference Number 17552). We also acknowledge support from the National Institutes of Health Research (http://www.nihr.ac.uk/Pages/default.aspx) Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London, from whom SDS receives funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ian.scott@kcl.ac.uk

Introduction

Rheumatoid arthritis (RA) is a common chronic inflammatory disorder. It results in substantial morbidity and disability alongside high medical and societal costs [1], [2]. There is therefore growing interest in preventing its development. Such prevention requires an ability to reliably predict who will develop RA. Advances in characterising genetic and environmental risk factors for RA together with developments in modelling methodology make predicting its development a realistic possibility.

RA is a clinical syndrome spanning multiple subsets [3]. The commonest subdivision is by the presence or absence of rheumatoid factor (RF)/anti-citrullinated protein antibodies (ACPA), termed seropositive and seronegative RA respectively.

Author Summary

Rheumatoid arthritis (RA) is a common, incurable disease with major individual and health service costs. Preventing its development is therefore an important goal. Being able to predict who will develop RA would allow researchers to look at ways to prevent it. Many factors have been found that increase someone's risk of RA. These are divided into genetic and environmental (such as smoking) factors. The risk of RA associated with each factor has previously been reported. Here, we demonstrate a method that combines these risk factors in a process called "prediction modelling" to estimate someone's lifetime risk of RA. We show that firstly, our prediction models can identify people with very high-risks of RA and secondly, they can be used to identify people at risk of developing RA at a younger age. Although these findings are an important first step towards preventing RA, as only a minority of people tested had substantially increased disease risks our models could not be used to screen the general population. Instead they need testing in people already at risk of RA such as relatives of affected patients. In this context they could identify enough numbers of high-risk people to allow preventive methods to be evaluated.

Risk factor evaluation has mainly focussed on seropositive RA with nearly half its genetic architecture known. *HLA-DRB1* alleles, in particular those encoding the shared epitope, dominate genetic risk accounting for approximately 36% of heritability [4]; 45 non-HLA variants explain approximately 15% of heritability [4]. Smoking is the main environmental risk factor [5]; it predisposes to seropositive RA and has a synergistic relationship with the shared epitope [6], [7]. Although single factors do not provide sufficient risk stratification, combining multiple factors within a prediction model may identify clinically relevant high- and low-risk groups. The large risks conferred by HLA make such modelling an attractive prospect in RA despite limited success in other complex disorders [8–10].

RA develops over many years prior to clinical presentation [11]. Initially, individuals with genetic susceptibility variants are exposed to environmental risks; some may develop autoantibodies (RF/ACPA) [12]. A proportion will subsequently develop arthralgia, which may progress to an unclassified arthritis followed by a fully expressed RA phenotype. Pilot studies in unclassified arthritis indicate that secondary prevention may be possible with corticosteroids [13], [14], methotrexate [15] and biologics [16] attenuating the progression to RA. Although preventive treatments may be more effective before immune dysregulation and symptoms develop, primary prevention is not currently possible as no reliable method exists to identify asymptomatic high-risk individuals.

Prevention is likely to have a larger impact in younger onset RA (YORA) due to the increased health costs associated with a longer disease duration [17]. Genetic susceptibility factors may influence RA's age of onset with *HLA-DRB1**04 alleles [18–21] and multiple single nucleotide polymorphisms (SNPs) such as those tagging *VEGFA* [22], *RANKL* [19], [23], *MMP1-3* [22] and *PTPN22* [24], [25] loci associating with YORA.

One group has published two reports outlining predictive models for RA. Their models, built using 8 HLA alleles, 14–31 SNPs and clinical factors, generated an aggregate weighted genetic risk score (wGRS) formed from the product of individual-locus odds ratios (ORs) [26], [27]. They were reasonably accurate at determining disease status in approximately 1,200 cases and 1,200 controls, with a maximal area under the curve (AUC) of 0.752. They also demonstrated a better ability to predict erosive RA (a more severe phenotype). However, only a minority of the studied populations had significantly elevated risks for RA.

We report an alternative modelling approach to predicting RA. Our novel modelling method uses computer simulation to categorise risk profiles; our models also incorporate a larger number of HLA risk variants. The risk factors included in our modelling comprise 15 four-digit/10 two-digit *HLA-DRB1* alleles, 31 SNPs and male ever-smoking status (as ever-smoking is a significant risk for RA in males only). We applied our models to two large cohorts of European ancestry: the Wellcome Trust Case Control Consortium (WTCCC) and the UK RA Genetics Group (UKRAGG) Consortium. Our primary aim was to determine if our approach would generate clinically relevant predictive values. Our secondary aim was to determine if our modelling better identified YORA. We demonstrate that clinically informative RA risk prediction is possible and that the risk of younger and older onset RA can be predicted using information on HLA and smoking status, respectively.

Materials and Methods

Ethics Statement

All participants in WTCCC and UKRAGG were recruited after providing informed consent. UKRAGG was approved by the North West Multi-Centre Research Ethics Committee (MREC 99/8/84). Authors gained written permission and approval from WTCCC to undertake this work in the publically available WTCCC1 collections.

Study Populations

The WTCCC dataset contains SNP data on 1,999 RA cases and 3,004 controls [28]. Controls were obtained from the 1958 British Birth Cohort and UK Blood Services. Genotyping was performed on the Affymetrix GeneChip 500k Mapping Array Set. Quality control (QC) procedures were undertaken excluding individuals with <97% SNP call rates, high heterozygosity, non-European ancestry or relatedness, discordance between genotype and phenotype data and duplicate samples. In the post-QC dataset information was available on 490,031 SNP markers; the total genotyping rate was 1.00. Two- or four-digit resolution *HLA-DRB1* tissue typing data were available on 1,837 cases and 1,647 controls.

The UKRAGG dataset contains SNP data on 5,024 RA cases and 4,281 controls from 6 UK centres [29]. Genotyping was performed using the Sequenom platform. Four hundred and four SNPs were genotyped over 8 staggered plexes; for each plex separate QC was undertaken excluding individuals and SNPs with <90% data present. In the post-QC dataset total genotyping rates were 0.73 owing to systematic differences in samples run on each plex. Two- or four-digit resolution *HLA-DRB1* tissue typing data were available on 3,420 cases and 1,500 controls.

Both datasets contained cases fulfilling the 1987 ACR classification criteria for RA [30]. *HLA-DRB1* tissue typing was undertaken (at two-digit or four-digit resolution) at individual centres, using commercially available semiautomated polymerase chain reaction-sequence-specific oligonucleotide probe (PCR-SSOP) typing techniques (or research assays based on PCR-SSOP linear array technology) [29]. Two-digit typing includes the allele group (Field 1) only; four-digit typing includes both the allele group and the allele subtype encoding a specific HLA protein (Field 2) (http://hla.alleles.org/nomenclature/naming.html).

Table 1. Clinical characteristics of WTCCC/UKRAGG cases and controls included in modelling.

		WTCCC		UKRAGG	
		RA (n = 1,516)	Controls (n = 1,647)	RA (n = 2,623)	Controls (n = 1,500)
Gender	Female	1,151 (76.0)	739 (50.0)	1,868 (71.2)	890 (59.9)
RA Characteristics	RF+	1,452 (96.1)	-	2,385 (93.1)	-
	ACPA+	1,061 (86.5)	-	1,508 (84.8)	-
	Mean Age Of Onset (95% CI)	45.3 (44.6–46.1)	-	48.0 (47.5–48.6)	-
	Erosive Disease	1,009 (71.1)	-	830 (69.7)	-
	Nodules	-	-	859 (38.3)	-
Smoking Status	Male Ever-Smokers	231 (80.5)[a]	422 (57.1)[a]	417 (78.8)[a]	149 (46.3)[a]
	Female Ever-Smokers	552 (58.3)[b]	425 (57.7)[b]	758 (55.9)[b]	238 (39.3)[b]

Data are number (%) unless otherwise stated. The following data are missing from WTCCC: gender in 2 cases and 169 controls; RF status in 5 cases; ACPA status in 290 cases; age of onset missing/inaccurate in 63 cases; erosive status in 96 cases; smoking status in 76 male cases, 204 female cases and 3 female controls. The following data are missing from UKRAGG: gender in 14 controls; RF status in 60 cases; ACPA status in 844 cases; age of onset missing/inaccurate in 93 cases; erosive status in 1,432 cases; nodular status in 378 cases; smoking status in 226 male cases, 513 female cases, 274 male controls and 284 female controls.
[a] = % of males that are ever smokers;
[b] = % of females that are ever smokers.
doi:10.1371/journal.pgen.1003808.t001

We undertook prediction modelling in seropositive cases and controls with *HLA-DRB1* tissue typing data available with or without additional SNP and smoking data (as most replicated risk loci are for seropositive RA and genetic risk is dominated by HLA) [4], [31]. The final cohorts comprised 1,516 cases and 1,647 controls from WTCCC and 2,623 cases and 1,500 controls from UKRAGG (Table 1).

Prediction Modelling Overview

Our modelling was performed within the R package, REGENT (Risk Estimation for Genetic and Environmental Traits), developed within our unit. This program incorporates published gene-environment risk factor and disease statistics to categorise risk using a confidence interval (CI)-based approach within a simulated population. The methodology underlying REGENT has previously been described in detail [32], [33].

Genetic and environmental risk factors for input into REGENT are selected from the literature. Genetic risk factors require allelic ORs, allele frequencies, and sample sizes from relevant studies, in order to estimate precision. Environmental risk factors require ORs, standard errors and the proportion of the population exposed to the risk factor. Data on these risk factors are entered into REGENT as summary statistic input files, which are processed in two stages: the first develops the prediction model and the second runs the prediction model in real life data.

In the first stage REGENT simulates a population-distribution of disease risk. Risk profiles are simulated based on the frequency of each risk factor in the general population. Summary ORs for each risk profile are generated through combining the ORs for each genetic and environmental risk factor in a multiplicative model that assumes risk factor independence. CIs are generated using information on the variability of genetic risk factors (derived from the sample size of the risk variant discovery cohort) and environmental risk factors (standard error of the effect size). Each simulated risk profile's OR is initially calculated relative to a profile with no risk factors present; these are subsequently adjusted to ensure correct disease prevalence in the population, assigning a risk profile with a mean OR as having a baseline risk of 1.0. CIs are used to classify risk profiles into four risk categories (reduced, average, elevated and high-risk). Starting with the risk profile of baseline risk (OR = 1.0), any risk profile whose CI overlaps with this baseline CI is classified as being of average-risk (as this profile is not statistically different from baseline). Any risk profile whose CI resides fully below the baseline CI is classified as reduced-risk. Profiles with CIs above the baseline CI are classified as elevated-risk. Furthermore, a high-risk group is determined by profiles whose CIs reside completely above the CI of the first risk profile classified as elevated-risk. An example of how this process is undertaken in a simplified model using 3 SNPs is provided in Figure S1.

In the second stage REGENT applies this simulated population profile to individual level data. Genotypes and environmental risk factor exposure data on each individual in the dataset of interest (WTCCC and UKRAGG) are entered into REGENT, which generates two measures of disease risk. Firstly, each individual's summary OR (95% CI) for RA is calculated (relative to the baseline individual with an OR of 1.0); as with the simulated population, risk factors are combined in a multiplicative model. This summary OR informs the individual of their risk of developing RA. Secondly, each individual is assigned a risk category for RA. This is undertaken through comparing the CI of each individual's summary OR to those of the simulated risk distribution in the same manner as described in stage 1. This risk category informs an individual whether they are at an increased or reduced risk of disease, relative to the average person in the general population.

Prediction Model Components Identified from Meta-Analyses

Genetic Risk Factors. We identified genetic susceptibility variants for potential inclusion in our prediction modelling from two large, recently published meta-analyses [34], [35]. We sought to include only susceptibility alleles attaining genome-wide significance ($P_{GWAS}<5\times10^{-8}$); this ensured that the alleles modelled were replicated RA genetic risk factors. These comprised 15 four-digit and 10 two-digit *HLA-DRB1* alleles and 35 non-HLA SNPs.

Environmental Risk Factor. We included the environmental risk factor smoking in our modelling. Other factors proposed to influence RA risk such as alcohol were not included: firstly the evidence underlying these is uncertain, with associations often present in case-control and not cohort studies [36] and secondly

detailed data on non-smoking risk factors were not captured in WTCCC and UKRAGG.

We used published ORs from the most recent meta-analysis evaluating smoking as an RA risk factor [5]. In this meta-analysis ever-smoking was a significant risk for seropositive RA in males only (OR 3.02; 95% CI 2.35–3.88) with a substantially smaller and non-significant (CIs contain 1) impact seen in females (OR 1.34; 95% CI 0.99–1.80). We therefore hypothesized that smoking would not improve prediction in women (confirmed in preliminary analyses; Table S1). As a result only males were evaluated in our modelling incorporating ever-smoking.

Although smoking interacts with the shared epitope we did not factor this into our modelling. This is because studies reporting summary ORs for this interaction [6], [29], [37], [38] have marked heterogeneity between them; therefore using meta-analysis techniques to obtain pooled ORs for shared epitope-smoking status combinations would be inaccurate and thus inappropriate. Examples of this heterogeneity include: (1) studies reporting risks stratified by different smoking levels, which would require an inverse variance fixed-effects model to obtain common ORs for all smokers within studies in addition to a random-effects model to estimate pooled ORs across studies; (2) two studies classifying the shared epitope at two-digit resolution, thus incorporating non-shared epitope alleles [6], [37]; (3) two studies not including all known shared epitope alleles [29], [38].

Prediction Model Component Availability in WTCCC and UKRAGG

Two-digit or four-digit *HLA-DRB1* tissue typing data were available in all evaluated individuals. In WTCCC 1,342 seropositive cases, 966 ACPA-positive cases and 1,126 controls had four-digit resolution data available on both alleles; 29 seropositive cases, 14 ACPA-positive cases and 159 controls had two-digit resolution data available on both alleles; 145 seropositive cases, 81 ACPA-positive cases and 362 controls had mixed-digit resolution data (one *HLA-DRB1* allele known at four-digit and the other at two-digit resolution) available. In UKRAGG 1,534 seropositive cases, 1,108 ACPA-positive cases and 735 controls had four-digit resolution data available on both alleles; 312 seropositive cases, 66 ACPA-positive cases and 205 controls had two-digit resolution data available on both alleles; 777 seropositive cases, 334 ACPA-positive cases and 560 controls had mixed-digit resolution data available.

We excluded 4 SNPs attaining P_{GWAS} in the meta-analysis for the following reasons: 1 (rs11676922) was in high linkage disequilibrium ($r^2>0.9$; HapMap release 22 CEU population panel) [39] with another (rs10865035) – in this case the latter SNP was included due to a previous association with RA – and 3 SNPs/proxy SNPs were unavailable (rs10488631, rs6859219 and rs934734 in UKRAGG; rs6822844, rs874040 and rs951005 in WTCCC). Eleven and two proxy SNPs were used in WTCCC and UKRAGG respectively (Table S2) [39].

Data on ever-smoking status were available in 287 male cases and 739 male controls in WTCCC and 529 male cases and 322 male controls in UKRAGG.

Final Prediction Models

To examine the contribution of each gene-environment component to prediction we constructed several models. These comprised a SNP model (with 31 SNPs), an HLA model (10 two-digit and 15 four-digit *HLA-DRB1* alleles), an HLA-SNP model (combining HLA and SNP model components), an HLA-smoking model (combining *HLA-DRB1* alleles with ever-smoking status) and an HLA-SNP-smoking model (combining *HLA-DRB1* alleles, 28 SNPs and ever-smoking status). Only the 28 SNPs present in both WTCCC and UKRAGG were incorporated in the last model. The latter two models, which included smoking, were evaluated in males only.

The decision to combine two-digit and four-digit *HLA-DRB1* alleles in the HLA model was undertaken to avoid removing the substantial number of individuals with mixed resolution typing. Preliminary analyses confirmed the validity of this approach with no significant differences seen in the discriminative abilities of HLA models incorporating (1) two-digit alleles only; (2) four-digit alleles only and (3) a mixed resolution of alleles (Table S3). Within our mixed resolution modelling the risks for each HLA allele were included only once per individual at the highest resolution at which they were known.

Only individuals with available data on relevant risk factors were included in models incorporating those risk factors. Therefore only males with available smoking data were included in the HLA-smoking and HLA-SNP-smoking models. Similarly only individuals with data available on the modelled SNPs could be included in the HLA-SNP and HLA-SNP-smoking models. Owing to missing data the number of individuals evaluated in each prediction model fell as more risk factors were included (Figure 1).

Statistical Analyses

Evaluating Dataset Validity. To compare the representativeness of our datasets to published RA populations we summarised clinical features of cases and controls (Table 1) and calculated effect allele frequencies and allelic ORs (95% CIs) (Tables 2 and 3). For the *HLA-DRB1* allele case-control association analysis (Table 2) the two-digit resolution allele results included both individuals with two-digit resolution typing and collapsed four-digit resolution typing. This approach was undertaken due to the small number of individuals with two-digit typing data in WTCCC/UKRAGG. The meta-analysis from which we obtained our risk alleles had almost identical allele frequencies when comparing two-digit alleles and four-digit alleles collapsed to two-digit resolution [35]; comparing our datasets to the meta-analysis findings in this manner was therefore appropriate.

Comparing Model Classification Abilities. To evaluate the ability of each model to correctly classify disease status we constructed receiver operating characteristic (ROC) curves and measured the AUC; this is established methodology in determining genetic classification test efficacy [40], [41]. Higher AUCs indicate better classification. An AUC>0.5 signifies some discriminative ability; a perfect classifier has an AUC of 1. AUCs were calculated and compared using DeLong's method [42] performed within the R package, pROC [43].

Comparing Model Generated Risk Distributions. The risk distributions for cases and controls under each model were compared by plotting the logarithmic OR for seropositive RA for each individual ordered by risk.

Calculating Lifetime Risk of RA. Due to the low prevalence of RA [44], ORs approximate relative risks [45]. Therefore to calculate lifetime risks of seropositive RA we multiplied published lifetime risks by the summary OR for RA generated by our prediction models. As UK lifetime risks of RA are unknown we used estimates from a large US cohort study (2.4% for women; 1.1% for men) [46].

Evaluating YORA Prediction. The role of HLA, SNPs and ever-smoking status in determining age of RA onset was evaluated using individual-level OR outputs from the REGENT models in a Cox univariate analysis with gender, smoking status and smoking status-gender interaction used as covariates. Factors indicated as likely predictors of age of onset were then examined

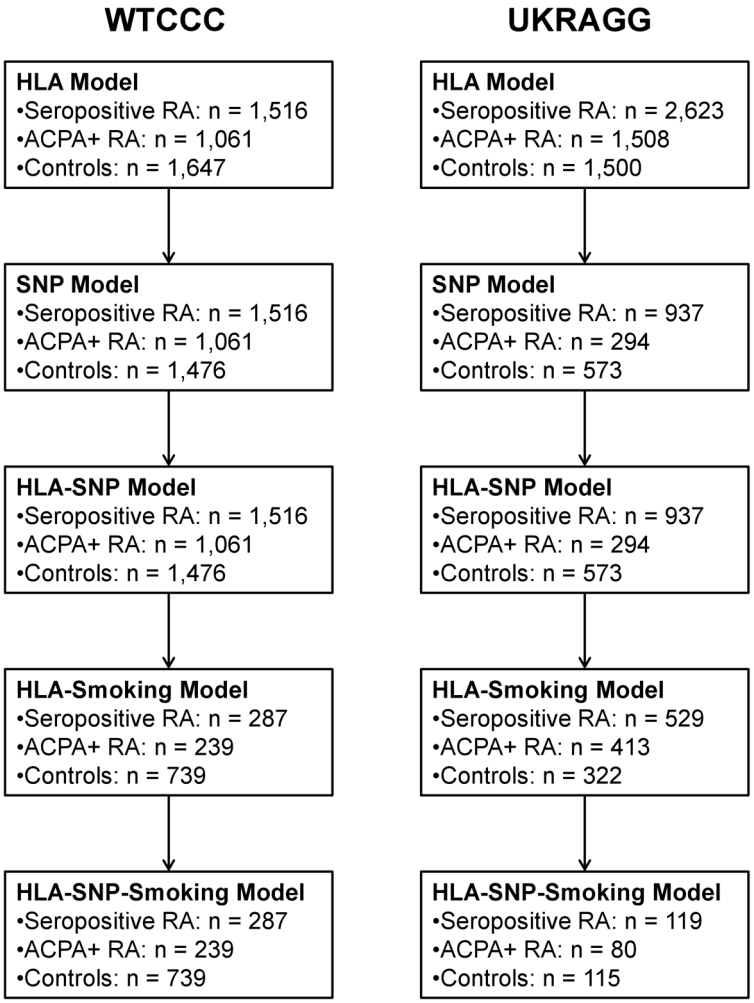


Figure 1. Number of individuals evaluated in each prediction model.
doi:10.1371/journal.pgen.1003808.g001

simultaneously in a multivariate analysis incorporating backward elimination of non-significant factors ($P>0.05$). We found no evidence of a "gender-smoking interaction" effect on the age of RA onset in either dataset (WTCCC $P=0.0823$ and UKRAGG $P=0.8369$; Table 4). This excluded a significant influence of gender on the relationship between smoking and the age at which RA developed. We therefore included both sexes when evaluating smoking's effect on the age of onset. Proportional hazards assumptions were verified using visual inspection of log-log plots [47]. To further demonstrate associations between significant factors and age of onset we constructed Kaplan-Meier estimates of the cumulative risk for cases, stratified by REGENT risk categorisation from the relevant models, alongside the presence/absence of other risk factors. We used a Cox multivariate approach to establish which four-digit *HLA-DRB1* alleles influenced age of onset (fitting all alleles simultaneously using stepwise selection, removing non-significant alleles from the final model). All time to event analyses were performed using SAS version 9.3 (SAS Institute, Cary, NC).

Separate Analyses for ACPA-Positive RA

We undertook modelling separately for seropositive (RF and/or ACPA present) RA and ACPA-positive RA since *HLA-DRB1* allelic ORs were obtained from a meta-analysis evaluating ACPA-positive RA [35], and the shared epitope alleles, non-HLA SNPs and smoking predominantly associate with ACPA-positive disease [4], [48–50]. We therefore hypothesised our modelling would perform better for ACPA-positive RA. As this was confirmed in the risk categorisation results we restricted further analyses (AUC and lifetime risk calculations, examining modelling associations with age of RA onset) to ACPA-positive RA.

Results

Dataset Validity

Genetic Risk Factors. In both WTCCC and UKRAGG the effect allele frequencies and ORs for seropositive RA were generally similar to published data (Tables 2 and 3). Exceptions occurred at the four-digit *HLA-DRB1* alleles *04:08 and *15:01 (absent from controls in WTCCC and UKRAGG respectively), at *01:01, *11:01, *11:04, *13:01 and *15:01 in WTCCC and *08:01 in UKRAGG (significantly lower allele frequencies in controls than expected). The absence of *04:08 in controls was probably a chance finding since it has a frequency of 0.005. The remaining discrepancies resulted from lower four-digit tissue typing rates for these alleles in controls, which were more often typed at two-digits, compared with cases. Although this could introduce bias,

Table 2. Classical *HLA-DRB1* allele frequencies and their association with seropositive RA in WTCCC and UKRAGG.

	Published Meta-Analysis [35]			WTCCC			UKRAGG		
HLA-DRB1 Allele	OR (95% CI)	MAF Co	MAF Ca	OR (95% CI)	MAF Co	MAF Ca	OR (95% CI)	MAF Co	MAF Ca
*01	1.30 (1.21–1.40)	0.113	0.145	1.53 (1.31–1.78)	0.104	0.151	1.27 (1.11–1.45)	0.121	0.149
*01:01	1.38 (1.28–1.50)	0.097	0.133	5.88 (4.62–7.55)	0.026	0.136	1.25 (1.06–1.47)	0.081	0.099
*03	0.59 (0.54–0.64)	0.128	0.082	0.67 (0.58–0.78)	0.148	0.105	0.76 (0.67–0.86)	0.159	0.125
*03:01	0.59 (0.54–0.64)	0.128	0.082	0.65 (0.55–0.76)	0.145	0.099	0.44 (0.37–0.51)	0.130	0.061
*04	3.71 (3.49–3.93)	0.174	0.450	2.90 (2.59–3.24)	0.213	0.439	3.19 (2.86–3.56)	0.184	0.419
*04:01	4.14 (3.86–4.44)	0.104	0.309	2.93 (2.57–3.35)	0.124	0.293	3.00 (2.63–3.42)	0.111	0.272
*04:04	3.17 (2.83–3.54)	0.036	0.091	1.86 (1.52–2.28)	0.052	0.092	2.56 (2.08–3.18)	0.039	0.093
*04:05	2.31 (1.77–3.01)	0.007	0.012	2.01 (1.12–3.73)	0.006	0.012	2.61 (1.34–5.58)	0.004	0.010
*04:08	5.48 (4.11–7.30)	0.005	0.017	-[a]	0.000	0.021	2.78 (1.70–4.76)	0.007	0.018
*07	0.49 (0.45–0.54)	0.133	0.064	0.48 (0.41–0.56)	0.154	0.080	0.54 (0.46–0.62)	0.142	0.081
*07:01	0.49 (0.45–0.54)	0.133	0.064	0.41 (0.35–0.49)	0.154	0.070	0.37 (0.32–0.44)	0.140	0.057
*08	0.41 (0.34–0.50)	0.029	0.013	0.39 (0.24–0.62)	0.022	0.009	0.30 (0.20–0.44)	0.029	0.009
*08:01	0.34 (0.26–0.44)	0.019	0.009	0.27 (0.13–0.53)	0.014	0.004	0.69 (0.33–1.46)	0.005	0.003
*10	2.53 (2.04–3.14)	0.008	0.020	1.97 (1.11–3.59)	0.006	0.012	1.75 (1.04–3.07)	0.007	0.012
*10:01	2.53 (2.04–3.14)	0.008	0.020	1.97 (1.11–3.59)	0.006	0.012	1.48 (0.85–2.67)	0.006	0.009
*11	0.48 (0.43–0.54)	0.094	0.039	0.50 (0.39–0.64)	0.064	0.033	0.42 (0.34–0.53)	0.065	0.028
*11:01	0.44 (0.38–0.52)	0.061	0.028	0.80 (0.55–1.14)	0.023	0.018	0.33 (0.23–0.47)	0.030	0.010
*11:04	0.15 (0.10–0.23)	0.024	0.008	0.79 (0.41–1.49)	0.008	0.006	0.38 (0.15–0.91)	0.005	0.002
*13	0.33 (0.30–0.37)	0.114	0.044	0.41 (0.33–0.50)	0.098	0.042	0.46 (0.38–0.55)	0.084	0.040
*13:01	0.28 (0.24–0.33)	0.061	0.021	0.77 (0.54–1.08)	0.026	0.020	0.42 (0.29–0.59)	0.027	0.011
*13:02	0.29 (0.23–0.38)	0.027	0.012	0.59 (0.38–0.90)	0.020	0.012	0.27 (0.18–0.42)	0.023	0.006
*14	0.50 (0.40–0.62)	0.025	0.012	0.51 (0.34–0.76)	0.024	0.013	0.45 (0.31–0.65)	0.023	0.010
*14:01	0.46 (0.36–0.59)	0.022	0.011	0.43 (0.28–0.66)	0.024	0.011	0.67 (0.34–1.33)	0.006	0.004
*15	0.59 (0.54–0.64)	0.142	0.092	0.63 (0.53–0.75)	0.128	0.084	0.60 (0.53–0.70)	0.146	0.093
*15:01	0.57 (0.53–0.62)	0.136	0.089	1.09 (0.87–1.37)	0.051	0.055	-[a]	0.000	0.025

All alleles attained genome-wide significance in the published meta-analysis; MAF = minor allele frequency; Co = controls; Ca = Cases;
[a] = OR incalculable due to no allele copies in the control group.
doi:10.1371/journal.pgen.1003808.t002

especially in the context of case-control association analyses, we do not consider it significantly affected our prediction modelling because these alleles were incorporated in our models at both two-digit and four-digit resolution (in most cases in the reference meta-analysis the two-digit alleles had similar allele frequencies and ORs compared with the four-digit alleles) and our risks were obtained from an external source [35].

SNP discrepancies occurred at rs3761847 in WTCCC and rs26232 and rs540386 in UKRAGG, which had ORs in the opposite direction to published results although the dataset and meta-analysis 95% CI's overlapped for two SNPs. Additionally the minor allele frequencies (MAFs) in controls were similar to those expected. These discrepancies probably represent normal variation as opposed to systematic genotyping differences.

Most *HLA-DRB1* alleles had significant associations with RA, with only 4 (16%) alleles in WTCCC and 3 (12%) alleles in UKRAGG having 95% CIs containing 1.0. A substantial proportion of SNPs – 13 (42%) in WTCCC and 15 (48%) in UKRAGG – had 95% CIs containing 1.0 reflecting their modest effect sizes, which required large discovery cohort sizes for detection.

Environmental Risk Factors. The ORs for seropositive RA in ever-smokers were 3.10 (95% CI 2.22–4.37) in WTCCC and 4.32 (95% CI 3.16–5.92) in UKRAGG for males and 1.02 (95% CI 0.84–1.25) in WTCCC and 1.96 (95% CI 1.61–2.40) in UKRAGG for females. The meta-analysis gender discrepancy surrounding the effect of ever-smoking on RA risk [5] was therefore mirrored in our datasets supporting the inclusion of only males in our smoking models.

Risk Prediction

Risk Categorisation. As hypothesized, our modelling more accurately categorised ACPA-positive RA as high-risk compared with seropositive RA (Figure 2 and Table S4). The HLA model provided most prediction in both datasets, classifying approximately one third of ACPA-positive RA as high-risk and two thirds of controls reduced-risk. Although the SNP model provided some prediction it classified most individuals as average-risk, reflecting the overlapping CIs generated by including many risk factors of a small effect size.

In WTCCC, the full genetic (HLA-SNP) model performed slightly better than HLA alone. Additional smoking data conferred subtle improvements in categorisation; this is particularly seen with the HLA-SNP-smoking model, which classified over half of ACPA-positive RA elevated/high-risk and 59% of controls reduced-risk.

In UKRAGG the addition of SNPs to HLA alleles increased the average-risk group size with no clear predictive benefits. The incorporation of smoking substantially improved prediction: the

Table 3. Non-HLA RA susceptibility SNP allele frequencies and their association with seropositive RA in WTCCC and UKRAGG.

	Published Meta-Analysis [34]			WTCCC		UKRAGG	
Loci	**SNP**	**MAF[a]**	**OR**	**MAF Ca/Co**	**OR (95% CI)**	**MAF Ca/Co**	**OR (95% CI)**
PTPN22	rs2476601	0.10	1.94 (1.81–2.08)	0.18/0.10	2.02 (1.73–2.36)	0.16/0.10	1.60 (1.38–1.85)
TNFAIP3	rs6920220	0.22	1.22 (1.16–1.29)	0.27/0.23	1.26 (1.12–1.41)	0.25/0.21	1.29 (1.15–1.44)
ANKRD55, IL6ST	rs6859219	0.21	0.78 (0.72–0.85)	0.17/0.20	0.80 (0.70–0.91)	-	-
CD40	rs4810485	0.25	0.85 (0.80–0.90)	0.22/0.24	0.87 (0.77–0.99)	0.22/0.25	0.83 (0.74–0.93)
CTLA4	rs3087243	0.44	0.87 (0.83–0.91)	0.43/0.44	0.95 (0.86–1.06)	0.43/0.47	0.86 (0.78–0.94)
TNFAIP3	rs5029937	0.04	1.40 (1.24–1.58)	0.06/0.04	1.58 (1.24–2.02)	0.05/0.04	1.39 (1.06–1.82)
IL2RA	rs706778	0.40	1.14 (1.09–1.20)	0.46/0.42	1.17 (1.05–1.29)	0.43/0.40	1.13 (1.02–1.25)
RBPJ	rs874040	0.30	1.14 (1.08–1.20)	-	-	0.33/0.31	1.11 (1.00–1.23)
TRAF1, C5	rs3761847	0.43	1.13 (1.08–1.18)	0.45/0.46	0.96 (0.87–1.07)	0.46/0.43	1.12 (1.01–1.24)
STAT4	rs7574865	0.22	1.16 (1.10–1.23)	0.21/0.19	1.12 (0.99–1.27)	0.25/0.22	1.18 (1.05–1.32)
SPRED2	rs934734	0.49	1.13 (1.08–1.19)	0.53/0.51	1.11 (1.00–1.23)	-	-
CCR6	rs3093023	0.43	1.13 (1.08–1.19)	0.42/0.40	1.10 (0.99–1.22)	0.47/0.44	1.16 (1.05–1.28)
PXK	rs13315591	0.09	1.29 (1.17–1.43)	0.10/0.09	1.11 (0.94–1.32)	0.08/0.07	1.10 (0.91–1.33)
C5orf30	rs26232	0.32	0.88 (0.84–0.93)	0.34/0.40	0.78 (0.70–0.86)	0.31/0.31	1.03 (0.92–1.14)
CCL21	rs951005	0.16	0.84 (0.78–0.90)	-	-	0.13/0.15	0.86 (0.75–0.99)
REL	rs13031237	0.37	1.13 (1.07–1.18)	0.45/0.43	1.07 (0.96–1.18)	0.41/0.37	1.22 (1.10–1.35)
AFF3	rs10865035	0.47	1.12 (1.07–1.17)	0.50/0.46	1.19 (1.07–1.31)	0.48/0.45	1.16 (1.05–1.27)
PRKCQ	rs4750316	0.19	0.87 (0.82–0.92)	0.16/0.20	0.77 (0.67–0.87)	0.18/0.19	0.89 (0.79–1.00)
IRF5	rs10488631	0.11	1.19 (1.10–1.28)	0.12/0.10	1.22 (1.04–1.44)	-	-
TNFRSF14	rs3890745	0.32	0.89 (0.85–0.94)	0.29/0.32	0.85 (0.76–0.95)	0.32/0.33	0.97 (0.88–1.08)
CD2, CD58	rs11586238	0.24	1.13 (1.07–1.19)	0.26/0.24	1.08 (0.96–1.21)	0.26/0.26	1.05 (0.93–1.17)
BLK	rs2736340	0.25	1.12 (1.07–1.18)	0.27/0.25	1.10 (0.98–1.24)	0.26/0.24	1.14 (1.01–1.28)
CD28	rs1980422	0.24	1.12 (1.06–1.18)	0.25/0.23	1.13 (1.01–1.28)	0.26/0.23	1.15 (1.03–1.30)
PRDM1	rs548234	0.33	1.10 (1.05–1.16)	0.36/0.34	1.11 (1.00–1.23)	0.35/0.35	1.00 (0.90–1.10)
CCL21	rs2812378	0.34	1.10 (1.05–1.16)	0.38/0.34	1.17 (1.05–1.30)	0.36/0.35	1.02 (0.92–1.13)
PTPRC	rs10919563	0.13	0.88 (0.82–0.94)	0.11/0.13	0.82 (0.70–0.95)	0.13/0.14	0.93 (0.80–1.07)
KIF5A, PIP4K2C	rs1678542	0.38	0.91 (0.87–0.96)	0.34/0.37	0.86 (0.77–0.95)	0.35/0.35	0.97 (0.88–1.07)
TRAF6	rs540386	0.14	0.88 (0.83–0.94)	0.11/0.13	0.90 (0.77–1.05)	0.14/0.13	1.03 (0.89–1.19)
FCGR2A	rs12746613	0.12	1.13 (1.06–1.21)	0.14/0.12	1.17 (1.01–1.36)	0.14/0.11	1.26 (1.08–1.46)
TAGAP	rs394581	0.30	0.91 (0.87–0.96)	0.28/0.30	0.92 (0.82–1.03)	0.28/0.29	0.94 (0.84–1.05)
TNFAIP3	rs10499194	0.27	0.91 (0.87–0.96)	0.25/0.27	0.90 (0.80–1.01)	0.26/0.28	0.90 (0.81–1.00)
IL2, IL21	rs6822844	0.18	0.90 (0.84–0.95)	-	-	0.15/0.19	0.80 (0.71–0.91)
IL2RA	rs2104286	0.27	0.92 (0.87–0.97)	0.24/0.27	0.85 (0.76–0.96)	0.25/0.26	0.94 (0.84–1.04)
IL2RB	rs3218253	0.26	1.09 (1.03–1.15)	0.29/0.25	1.22 (1.09–1.37)	0.29/0.27	1.09 (0.98–1.22)

SNPs are ordered by significance (most significant by P_{GWAS} listed first); all alleles attained genome-wide significance in the published meta-analysis; Ca = Cases; Co = Controls; MAF = Minor Allele Frequency;
[a] = MAF in controls.
doi:10.1371/journal.pgen.1003808.t003

HLA-SNP-smoking model classified 38% ACPA-positive RA vs. 3% controls as high-risk and 70% controls vs. 18% ACPA-positive RA as reduced-risk.

The general trend of improved prediction through modelling increasing numbers of risk factors is highlighted by the ratios of the percentage of ACPA-positive cases to controls classified high-risk by each model. In WTCCC these comprise 3.4 for the SNP model, 3.8 for the HLA model, 5.8 for the HLA-SNP model, 4.8 for the HLA-smoking model and 6.0 for the HLA-SNP-smoking model. Similarly, the ratios of the percentage of controls to ACPA-positive cases classified reduced-risk in WTCCC comprise 2.3 for the SNP model, 2.4 for the HLA model, 3.4 for the HLA-SNP model, 4.0 for the HLA-smoking model and 4.7 for the HLA-SNP-smoking model. Similar findings were present in UKRAGG.

AUC Assessments. In WTCCC AUCs for the SNP, HLA, HLA-SNP, HLA-smoking and HLA-SNP-smoking models in discriminating between ACPA-positive RA and controls comprised 0.660 (95% CI 0.638–0.681), 0.764 (95% CI 0.746–0.782), 0.796 (95% CI 0.779–0.813), 0.813 (95% CI 0.784–0.841) and 0.837 (95% CI 0.810–0.865), respectively (Figure 3). Significant differences in AUCs were observed between all three genetic models: SNP and HLA models $P<0.0001$; HLA and HLA-SNP models $P=0.0118$. Smoking data significantly improved discrim-

Table 4. Relationship between modelling components and age of RA onset.

	WTCCC			UKRAGG				
		Univariate Analysis[c]			Univariate Analysis		Multivariate Analysis	
Modelling Component	No. Cases Examined	*P*-Value	Hazard Ratio (95% CI)	No. Cases Examined	*P*-Value	Hazard Ratio (95% CI)	*P*-Value	Hazard ratio (95% CI)
HLA[a,b]	1022	<0.0001	1.034 (1.018–1.050)	1456	0.0004	1.025 (1.011–1.038)	0.0003	1.026 (1.012–1.040)
SNP[a]	1022	0.1804	1.043 (0.981–1.110)	284	0.294	1.075 (0.939–1.230)	-	-
Gender[b]	1021	0.2157	0.914 (0.792–1.054)	1456	0.0107	0.864 (0.722–0.967)	0.0465	0.885 (0.786–0.998)
Smoking[b]	962	0.1301	0.902 (0.789–1.031)	1361	0.0009	0.830 (0.743–0.927)	0.0041	0.848 (0.757–0.949)
Gender-Smoking Interaction[b]	961	0.0823	0.870 (0.744–1.018)	1361	0.009	0.846 (0.746–0.959)	0.8369	-

[a] = HLA and SNP variables represent the summary OR scores generated by the models incorporating HLA and SNP data respectively;
[b] = variables included in UKRAGG multivariate model after variable pruning using backwards selection and model comparison with Akaike's Information Criterion;
[c] = as only one parameter was significant in the WTCCC univariate analysis no multivariate model was fitted.
doi:10.1371/journal.pgen.1003808.t004

ination with differences observed between HLA and HLA-smoking model AUCs ($P = 0.0051$) and HLA-SNP and HLA-SNP-smoking model AUCs ($P = 0.0120$).

In UKRAGG AUCs for the SNP, HLA, HLA-SNP, HLA-smoking and HLA-SNP-smoking models in discriminating between ACPA-positive RA and controls comprised 0.617 (95% CI 0.577–0.656), 0.748 (95% CI 0.731–0.765), 0.756 (95% CI 0.723–0.790), 0.813 (95% CI 0.782–0.845) and 0.857 (95% CI 0.804–0.910), respectively (Figure 3). The HLA model had significantly better discrimination than the SNP model ($P<0.0001$). Combined SNP and HLA data did not improve discrimination with no differences observed between AUCs for the HLA and HLA-SNP models ($P = 0.665$) or the HLA-smoking and HLA-SNP-smoking models ($P = 0.1671$). Additional smoking information significantly improved modelling discrimination with significant differences observed between HLA and HLA-smoking model AUCs ($P = 0.0003$).

An overview of the main findings for each of the 5 prediction models, alongside the differences between them is provided in Figure S2.

Risk Distributions. In both datasets the HLA model provided most risk prediction generating substantially higher and lower ORs for RA in cases and controls respectively compared with the SNP model (Figure 4).

In WTCCC the addition of other risk factors to the *HLA-DRB1* alleles resulted in further small incremental increases in ORs for RA in cases; a less pronounced reduction in risk was seen in controls.

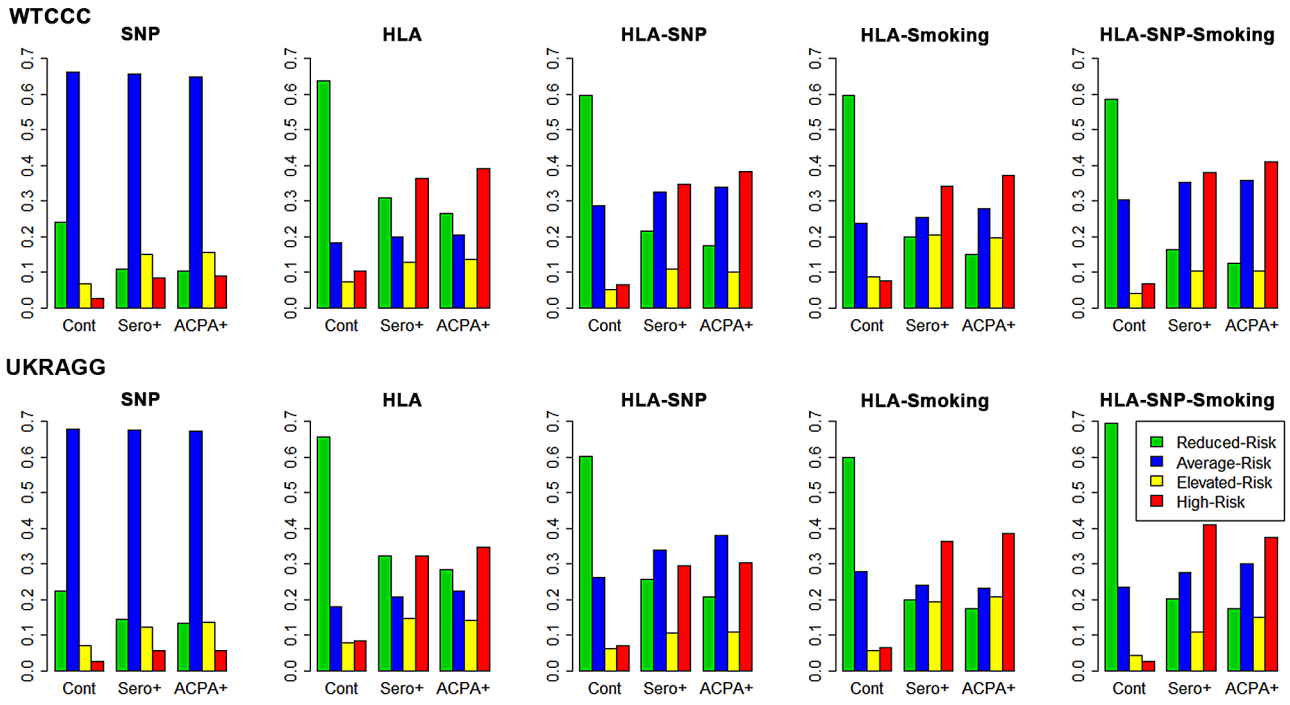


Figure 2. Risk categorisation of RA and controls by each prediction model. The y-axis on each graph refers to the proportion of cases/controls in each risk category; cont = controls; sero+ = seropositive RA; ACPA+ = ACPA-positive RA.
doi:10.1371/journal.pgen.1003808.g002

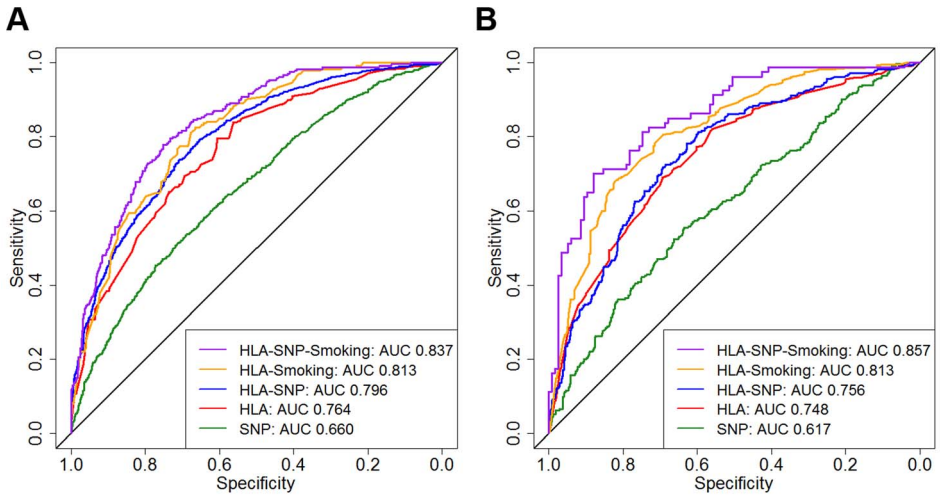


Figure 3. Prediction model receiver operating characteristic curves. Panel A = WTCCC; Panel B = UKRAGG; ROCs calculated for discriminating between ACPA-positive RA and controls; AUC = area under the curve. WTCCC model AUC comparisons: SNP versus HLA, $P<0.0001$; HLA versus HLA-SNP, $P=0.0118$; HLA-SNP versus HLA-Smoking, $P=0.3327$; HLA-Smoking versus HLA-SNP-Smoking, $P=0.0001$. UKRAGG model AUC comparisons: SNP versus HLA, $P<0.0001$; HLA versus HLA-SNP, $P=0.665$; HLA-SNP versus HLA-Smoking, $P=0.0145$; HLA-Smoking versus HLA-SNP-Smoking, $P=0.1671$.
doi:10.1371/journal.pgen.1003808.g003

In UKRAGG the addition of SNPs to HLA data provided no changes in case risk profiles, although a minority of controls had lower ORs. Additional smoking data resulted in significantly higher ORs for cases; only the HLA-SNP-smoking model clearly generated lower risk profiles for controls.

Lifetime Risk Prediction. Evaluating risks using genetics (HLA-SNP model) alone the highest risk WTCCC ACPA-positive case had an OR for seropositive RA of 79; as a male his lifetime risk was estimated at 86%. The highest risk control had an OR of 22; as a female her lifetime risk was estimated at 53%. Despite such high individual odds only a relative minority had relevant increased lifetime risks: using the same HLA-SNP model 49 (4.61%) ACPA-positive cases and 1 (0.07%) control had ORs for seropositive RA>20 (lifetime risks >48% if female and >22% if male) in WTCCC. In UKRAGG 9 (3.06%) ACPA-positive cases and 1 (0.17%) control had ORs>20.

The HLA-SNP-smoking model identified the greatest proportion of cases with substantially increased lifetime risks for RA. This model identified 18 (7.53%) and 3 (3.75%) ACPA-positive male cases to have ORs for seropositive RA>20 (lifetime risk >22%) in WTCCC and UKRAGG respectively; no controls had ORs>20.

Younger Onset RA Prediction

In WTCCC the HLA model summary OR score was the only significant predictor of age of RA onset (Table 4). The hazard ratio (HR) was 1.034 ($P<0.0001$), which indicated that the hazard (the rate at which RA occurred) was greater in individuals with higher HLA derived ORs than those with lower ORs. Therefore a higher HLA model generated risk score associated with RA occurring at a faster rate and thus YORA. Conversely ever-smoking was associated with older onset RA: the HR of 0.902 indicated a smaller hazard (RA occurred at a slower rate) in ever-smokers compared with never-smokers, although this was not significant ($P=0.1301$).

In UKRAGG the HLA model summary OR score, gender and smoking status were significant independent predictors of age of

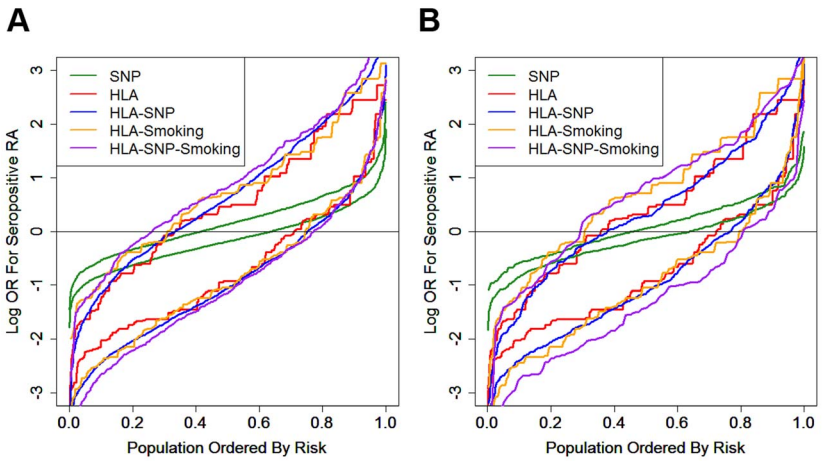


Figure 4. Prediction model generated risk profiles for ACPA-positive RA and controls. Panel A = WTCCC; Panel B = UKRAGG; the upper set of lines for each model refer to RA cases; the lower set of lines refer to controls; OR = odds ratio.
doi:10.1371/journal.pgen.1003808.g004

onset. An increasing HLA summary OR score associated with YORA ($P = 0.0003$, HR 1.026); ever-smoking ($P = 0.0041$, HR 0.848) and male gender ($P = 0.0465$, HR 0.885) associated with older onset RA.

We considered that the non-significant relationship between smoking and age of onset in WTCCC reflected a limited sample size with our power to detect a 0.88 HR in the 962 WTCCC cases approximately 51% compared with 65% for the 1,361 UKRAGG cases. We therefore undertook a pooled analysis of both datasets (incorporating an additional "study" variable to account for dataset median age of onset differences). This confirmed that HLA derived risk scores significantly associated with YORA ($P<0.0001$, HR 1.030) and ever-smoking significantly associated with older onset RA ($P = 0.0489$, HR 0.889).

Kaplan-Meier curves of age of onset stratified by HLA model risk categorisation further demonstrate the association of HLA risk profiles with YORA (Figure 5) with cases classified high-risk having significantly younger onset ages compared to those classified reduced-risk. In WTCCC the difference in the median time to RA (time point at which half the cases have developed RA) was 3 years between those classed high- and reduced-risk (Log-Rank = 11.43; $P = 0.0007$). In UKRAGG a stronger association was seen (Log-Rank = 27.33; $P<0.0001$) with a difference in median time to RA onset between risk groups of 6 years. Further stratification by ever-smoking status demonstrated a trend towards an older onset age in ever-smokers. In WTCCC the median time to onset difference between high-risk never-smokers and reduced-risk ever-smokers was 7 years (Log-Rank = 14.42; $P = 0.0024$); a larger disparity was seen in UKRAGG with a difference of 12 years observed (Log-Rank = 46.2505; $P<0.0001$).

Examining which four-digit resolution *HLA-DRB1* alleles influenced onset age revealed significant associations between age of onset and *03:01 ($P = 0.0313$), *04:01 ($P = 0.0001$), *04:08 ($P = 0.0032$) and *13:02 ($P = 0.0097$) in WTCCC and *04:01 ($P<0.0001$) and *04:04 ($P = 0.0243$) in UKRAGG. Three of these alleles (*04:01, *04:04 and *04:08) are shared-epitope alleles.

Discussion

We have demonstrated that predicting RA development is possible with our prediction models able to identify individuals with clinically relevant increased risks for seropositive RA. Our modelling indicates that most prediction is provided by *HLA-DRB1* alleles and, to a lesser extent, smoking in males; non-HLA susceptibility SNPs provide only minor predictive benefits. These findings are consistent with the estimations of heritability variance

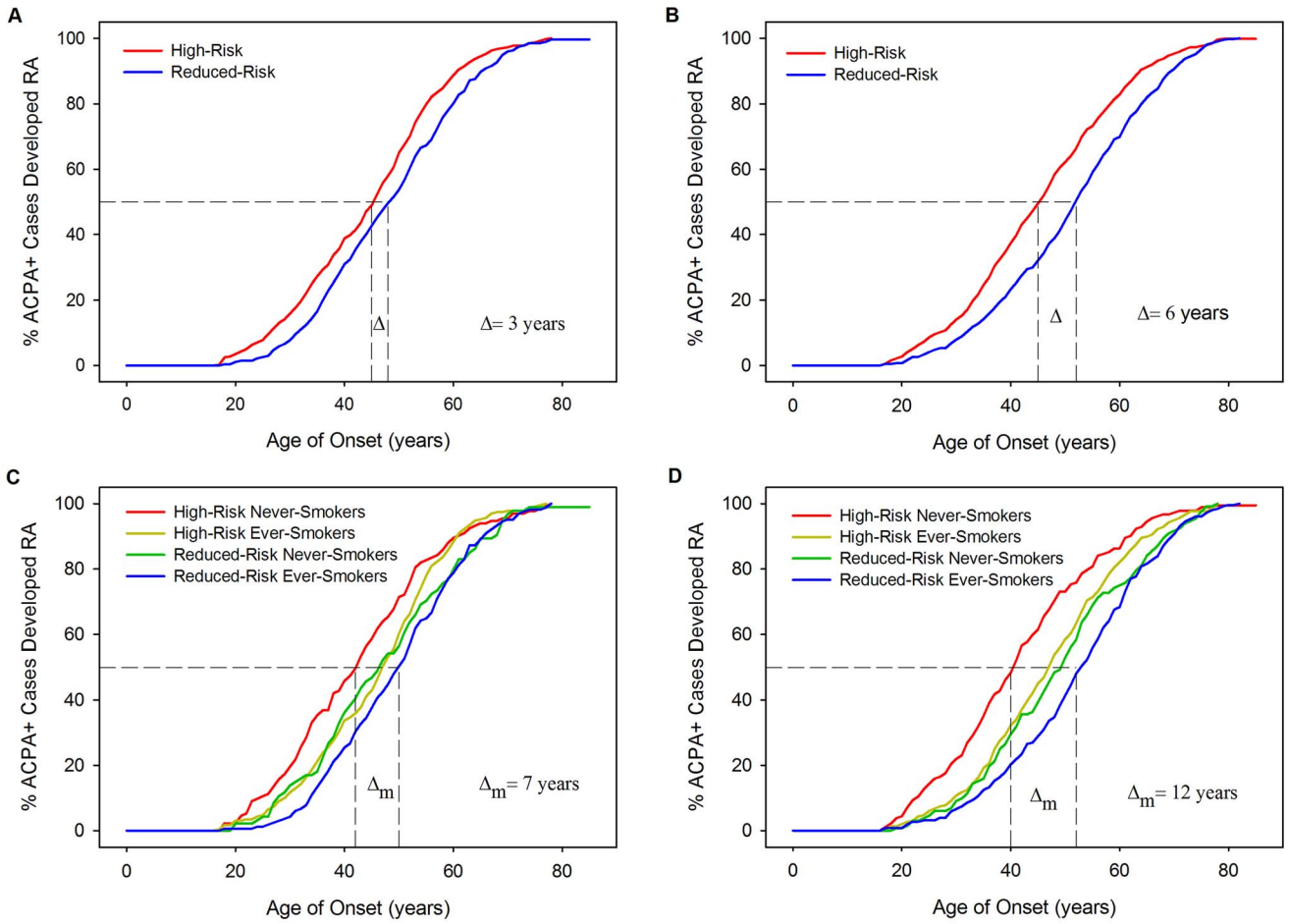


Figure 5. Kaplan-Meier curves: RA age of onset stratified by HLA model risk categorisation and smoking status. Panel A = WTCCC Curves Stratified By Risk Categorisation; Panel B = UKRAGG Curves Stratified By Risk Categorisation; Panel C = WTCCC Curves Stratified By Risk Categorisation and Ever-Smoking Status; Panel D = UKRAGG Curves Stratified By Risk Categorisation and Ever-Smoking Status; Δ = change in onset age; Δ_m = maximum change in onset age across strata.
doi:10.1371/journal.pgen.1003808.g005

conferred by different genetic components. We have also shown it is possible to predict the age of RA onset, using information on HLA and smoking to identify those at risk of younger and older onset RA, respectively. Whilst our novel modelling approach, which uses computer simulation-based categorisation alongside a greater number of HLA alleles, significantly improves upon the discriminative abilities of existing models [26], [27] it remains unsuitable for population screening with only a minority at significantly increased lifetime risks for RA.

Our approach provides some potential advantages over existing RA prediction modelling [26], [27]. Firstly, by using a simulated population to generate risk profiles we do not require an entire population of real-life data to stratify risks. In contrast existing approaches categorise wGRS scores using their Gaussian distribution in control groups. Secondly, our CI-based approach considers the precision with which risk factor effect sizes are known when classifying risk; this prevents classifying people high-risk if their risk is imprecisely known. Thirdly, our models provide greater discrimination: the highest AUC for existing clinical-genetic models in discerning ACPA-positive RA from controls is 0.752; the highest AUC for our clinical-genetic model is 0.857.

SNPs provided only minor improvements in prediction, highlighting the limitations of genome-wide association study (GWAS) derived data in this field. Although GWAS-established SNPs have helped identify cellular pathways relevant to RA pathogenesis [51] their modest effect sizes limit their predictive utility. It has been proposed that the missing heritability of RA may reflect the involvement of rare variants of large effect sizes or structural variants [52]. Alternative genotyping technologies such as next-generation sequencing may identify these variants, although only loci with large effect sizes will substantially improve prediction modelling.

Although individuals with clinically relevant increased lifetime risks (such as 86%) for RA were identified there was, overall, only a minority of individuals at a significantly elevated risk: 7% of ACPA-positive individuals had lifetime risks of 22% or more when evaluated using all available risk factors. Therefore despite high AUCs our modelling is unsuitable for population level screening. However, if its use was targeted to groups with *a priori* increased risks, such as first degree relatives of RA probands [53–55], then a substantially greater proportion of very high-risk individuals might be identified.

Individuals classified high-risk by our HLA model were more likely to develop RA at a younger age. This finding – mainly attributable to the *04:01 allele – is supported by existing literature. Hellier *et al* reported a higher frequency of *04 RA associated alleles in YORA (present in 52% of 262 RA cases with onset age <60) compared with elderly onset RA (present in 37% of 60 cases with onset age >60; $P = 0.045$) [18]. Similarly, Wu *et al* identified a significantly younger age of onset in Caucasian RA patients carrying shared epitope encoding *04 alleles ($P = 0.0003$) [19]. Other studies report positive correlations between YORA and shared epitope alleles [25], [56]. Our finding of ever-smoking associating with older onset RA is less established. It has only been examined in three relatively small studies, with contrasting outcomes: one study reported a significant relationship between smoking at disease onset and a younger onset age [57]; one reported a younger onset age in current vs. never-smokers (although ex-smokers had older onset RA in comparison to both these groups) [58]; the final study found no association [59]. Our findings – demonstrated in 2,323 individuals across two independent datasets – are biologically plausible. As risk genotypes are present from birth they can exert their effects on disease risk throughout an individual's lifetime; therefore possessing high-risk *HLA-DRB1* alleles predisposes to RA at a younger age. In contrast the risk of RA increases as more cigarettes are smoked [60], [61] and smokers are exposed to more cigarettes as they age; therefore smokers are more likely to develop RA as they get older because they have been exposed to more cigarettes and thus smoking associates with older onset RA. This logic also explains why ever-smoking associates with older onset RA in both men and women, with heavy smoking a risk factor for RA in both genders [5]. We were, however, unable to incorporate heavy smoking in our prediction modelling due to a paucity of data on smoking pack-years in WTCCC/UKRAGG.

We incorporated many genetic risk factors in our modelling but included only one environmental risk factor, smoking. This reflects uncertainty regarding relevant environmental risks alongside limited environmental data within current genetic datasets. Although many environmental factors are linked to RA their associations are usually identified in case-control studies, which are subject to multiple biases, rather than cohort studies. Examples include alcohol consumption [36], parity [62], [63] and oral contraceptive pill use [64]. Better characterisation of environmental risks will enhance predictive modelling.

Our modelling has several limitations. Firstly, WTCCC participants were included in the meta-analyses that we obtained our genetic risk loci data from; however WTCCC comprised only a proportion of the meta-analyses datasets (20% of the HLA meta-analysis; 29% of the SNP meta-analysis) and our findings were independently replicated in UKRAGG. Secondly, missing data meant the number of individuals included in each model fell as more risk factors were included; this is particularly seen in models incorporating smoking. Thirdly, due to marked heterogeneity in published data on gene-gene/gene-environment interactions we assumed independence between these factors despite known interactions existing between the shared epitope alleles and *PTPN22* and smoking [6], [7], [29], [37], [38].

Improving RA prediction requires better clarification of its genetic and environmental risk factors. Identifying risk factors with large effect sizes of known precision will most enhance prediction modelling. This could be facilitated through fine-mapping studies that better tag causal variants [65] alongside prospective cohort studies examining environmental risk factors in RA cases subdivided by ACPA status, with increasing evidence that risks differ between these serological subsets [36], [66]. It is, however, unlikely that identifying such risk factors will substantially increase the proportion of individuals with clinically relevant high disease risks. We therefore consider that prediction modelling requires evaluation in *a priori* higher risk groups. In this context it may identify sufficient numbers of very high-risk individuals, facilitating a better understanding of pre-RA immunopathology and enabling the assessment of primary prevention strategies.

Supporting Information

Figure S1 REGENT stage 1- simulation of general population risk profiles: example using a model incorporating 3 SNPs.
(TIFF)

Figure S2 Main findings of each prediction model.
(TIFF)

Table S1 Risk categorisation results for models incorporating smoking in females. Data are number (%) unless stated otherwise; [a] = AUCs calculated using ACPA-positive cases.
(DOCX)

Table S2 Proxy SNPs used in modelling. a = proxy SNP obtained using 1,000 Genomes CEU population panel [39]; b = proxy SNP obtained using HapMap release 22 CEU population panel [39]; c = proxy SNP obtained using Ricopili (Broad Institute, Boston, USA) from the GWAS meta-analysis of RA risk (http://www.broadinstitute.org/mpg/ricopili/).
(DOCX)

Table S3 Two-digit, four-digit and mixed-digit hla prediction model results, showing similar discriminative abilities. Data are number (%) unless stated otherwise; Sero+ = seropositive RA; ACPA+ = ACPA-positive RA; AUCs calculated using ACPA-positive cases; two-digit model evaluated individuals with all available HLA data collapsed down to two-digit resolution.
(DOCX)

Table S4 Prediction model results: risk categorisation. Data are number (%); Sero+ = seropositive RA; ACPA+ = ACPA-positive RA.
(DOCX)

Acknowledgments

We are grateful to all patients and staff contributing to WTCCC and UKRAGG.

Author Contributions

Conceived and designed the experiments: ICS SDS SS APC CML. Performed the experiments: ICS SDS RT. Analyzed the data: ICS SDS RT PF CML. Contributed reagents/materials/analysis tools: SS AH SE AWM AGW LJH PW AB JW. Wrote the paper: ICS SDS SS APC CML.

References

1. Wolfe F, Hawley DJ (1998) The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. J Rheumatol 25: 2108–2117.
2. Michaud K, Messer J, Choi HK, Wolfe F (2003) Direct medical costs and their predictors in patients with rheumatoid arthritis: a three-year study of 7,527 patients. Arthritis Rheum 48: 2750–2762.
3. Scott DL, Wolfe F, Huizinga TWJ (2010) Rheumatoid arthritis. Lancet 376: 1094–1108.
4. Eyre S, Bowes J, Diogo D, Lee A, Barton A, et al. (2012) High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 44: 1336–1340.
5. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, et al. (2010) Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 69: 70–81.
6. Pedersen M, Jacobsen S, Garred P, Madsen HO, Klarlund M, et al. (2007) Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. Arthritis Rheum 56: 1446–1453.
7. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, et al. (2007) Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet 80: 867–875.
8. Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, et al. (2009) Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. Diabetes Care 32: 2116–2119.
9. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, et al. (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 340: b4838.
10. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, et al. (2010) Association between a literature-based genetic risk score and cardiovascular events in women. JAMA 303: 631–637.
11. Gerlag DM, Raza K, van Baarsen LGM, Brouwer E, Buckley CD, et al. (2012) EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Ann Rheum Dis 71: 638–641.
12. Nielen MMJ, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, et al. (2004) Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum 50: 380–386.
13. Bos WH, Dijkmans BAC, Boers M, van de Stadt RJ, van Schaardenburg D (2010) Effect of dexamethasone on autoantibody levels and arthritis development in patients with arthralgia: a randomised trial. Ann Rheum Dis 69: 571–574.
14. Verstappen SM, McCoy MJ, Roberts C, Dale NE, Hassell AB, et al. (2010) Beneficial effects of a 3-week course of intramuscular glucocorticoid injections in patients with very early inflammatory polyarthritis: results of the STIVEA trial. Ann Rheum Dis 69: 503–509.
15. van Dongen H, van Aken J, Lard LR, Visser K, Ronday HK, et al. (2007) Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. Arthritis Rheum 56: 1424–1432.
16. Emery P, Durez P, Dougados M, Legerton CW, Becker JC, et al. (2010) Impact of T-cell costimulation modulation in patients with undifferentiated inflammatory arthritis or very early rheumatoid arthritis: a clinical and imaging study of abatacept (the ADJUST trial). Ann Rheum Dis 69: 510–516.
17. Lajas C, Abasolo L, Bellajdel B, Hernandez-Garcia C, Carmona L, et al. (2003) Costs and predictors of costs in rheumatoid arthritis: a prevalence-based study. Arthritis Rheum 49: 64–70.
18. Hellier JP, Eliaou JF, Daures JP, Sany J, Combe B (2001) HLA-DRB1 genes and patients with late onset rheumatoid arthritis. Ann Rheum Dis 60: 531–533.
19. Wu H, Khanna D, Park G, Gersuk V, Nepom GT, et al. (2004) Interaction between RANKL and HLA-DRB1 genotypes may contribute to younger age at onset of seropositive rheumatoid arthritis in an inception cohort. Arthritis Rheum 50: 3093–3103.
20. Jaraquemada D, Ollier W, Awad J, Young A, Silman A, et al. (1986) HLA and rheumatoid arthritis: a combined analysis of 440 British patients. Ann Rheum Dis 45: 627–636.
21. MacGregor A, Ollier W, Thomson W, Jawaheer D, Silman A (1995) HLA-DRB1*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. J Rheumatol 22: 1032–1036.
22. Chen Y, Mattey DL (2012) Age at onset of rheumatoid arthritis: association with polymorphisms in the vascular endothelial growth factor A(VEGFA) gene and an intergenic locus between matrix metalloproteinase (MMP) 1 and 3 genes. Clin Exp Rheumatol 30: 894–898.
23. Tan W, Wu H, Zhao J, Derber LA, Lee DM, et al. (2010) A functional RANKL polymorphism associated with younger age at onset of rheumatoid arthritis. Arthritis Rheum 62: 2864–2875.
24. Steer S, Lad B, Grumley JA, Kingsley GH, Fisher SA (2005) Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: evidence for an early onset/disease severity effect. Arthritis Rheum 52: 358–360.
25. Karlson EW, Chibnik LB, Cui J, Plenge RM, Glass RJ, et al. (2008) Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study. Ann Rheum Dis 67: 358–363.
26. Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, et al. (2010) Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann Rheum Dis 69: 1077–1085.
27. Chibnik LB, Keenan BT, Cui J, Liao KP, Costenbader KH, et al. (2011) Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. PLoS ONE [Electronic Resource] 6: e24380.
28. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
29. Morgan AW, Thomson W, Martin SG, Yorkshire Early Arthritis Register Consortium, Carter AM, et al. (2009) Reevaluation of the interaction between HLA-DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. Arthritis Rheum 60: 2565–2576.
30. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324.
31. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, et al. (2000) Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 43: 30–37.
32. Goddard GHM, Lewis CM (2010) Risk categorization for complex disorders according to genotype relative risk and precision in parameter estimates. Genet Epidemiol 34: 624–632.
33. Crouch DJ, Goddard GH, Lewis CM (2013) REGENT: a risk assessment and classification algorithm for genetic and environmental factors. Eur J Hum Genet 21: 109–111.
34. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514.
35. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee H-S, et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 44: 291–296.
36. Scott IC, Tan R, Stahl D, Steer S, Lewis CM, et al. (2013) The Protective Effect Of Alcohol On Developing Rheumatoid Arthritis: A Systematic Review And Meta-Analysis. Rheumatology (Oxford) [Epub ahead of print].
37. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, et al. (2011) Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. Ann Rheum Dis 70: 508–511.

38. Karlson EW, Chang SC, Cui J, Chibnik LB, Fraser PA, et al. (2010) Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. Ann Rheum Dis 69: 54–60.
39. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24: 2938–2939.
40. Metz CE (1978) Basic principles of ROC analysis. Seminars in Nuclear Medicine 8: 283–298.
41. Lu Q, Elston RC (2008) Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes. Am J Hum Genet 82: 641–651.
42. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44: 837–845.
43. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12: 77.
44. Symmons D, Turner G, Webb R, Asten P, Barrett E, et al. (2002) The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford) 41: 793–800.
45. Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 280: 1690–1691.
46. Crowson CS, Matteson EL, Myasoedova E, Michet CJ, Ernste FC, et al. (2011) The lifetime risk of adult-onset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum 63: 633–639.
47. Hess KR (1995) Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Stat Med 14: 1707–1723.
48. Huizinga TWJ, Amos CI, van der Helm-van Mil AHM, Chen W, van Gaalen FA, et al. (2005) Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 52: 3433–3438.
49. Orozco G, Pascual-Salcedo D, Lopez-Nevot MA, Cobo T, Cabezon A, et al. (2008) Auto-antibodies, HLA and PTPN22: susceptibility markers for rheumatoid arthritis. Rheumatology (Oxford) 47: 138–141.
50. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, et al. (2006) A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 54: 38–46.
51. Plenge RM (2009) Recent progress in rheumatoid arthritis genetics: one step towards improved patient care. Curr Opin Rheumatol 21: 262–271.
52. de Vries RRP, van der Woude D, Houwing JJ, Toes REM (2011) Genetics of ACPA-positive rheumatoid arthritis: the beginning of the end? Ann Rheum Dis 70 Suppl 1: i51–54.
53. Wolfe F, Kleinheksel SM, Khan MA (1988) Prevalence of familial occurrence in patients with rheumatoid arthritis. Br J Rheumatol 27 Suppl 2: 150–152.
54. Deighton CM, Wentzel J, Cavanagh G, Roberts DF, Walker DJ (1992) Contribution of inherited factors to rheumatoid arthritis. Ann Rheum Dis 51: 182–185.
55. Koumantaki Y, Giziaki E, Linos A, Kontomerkos A, Kaklamanis P, et al. (1997) Family history as a risk factor for rheumatoid arthritis: a case-control study. J Rheumatol 24: 1522–1526.
56. Fries JF, Wolfe F, Apple R, Erlich H, Bugawan T, et al. (2002) HLA-DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: frequency, severity, and treatment bias. Arthritis Rheum 46: 2320–2329.
57. Hutchinson D, Shepstone L, Moots R, Lear JT, Lynch MP (2001) Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. Ann Rheum Dis 60: 223–227.
58. Papadopoulos NG, Alamanos Y, Voulgari PV, Epagelis EK, Tsifetaki N, et al. (2005) Does cigarette smoking influence disease expression, activity and severity in early rheumatoid arthritis patients? Clin Exp Rheumatol 23: 861–866.
59. Diaz FJ, Rojas-Villarraga A, Salazar JC, Iglesias-Gamarra A, Mantilla RD, et al. (2011) Anti-CCP antibodies are associated with early age at onset in patients with rheumatoid arthritis. Joint, Bone, Spine: Revue du Rhumatisme 78: 175–178.
60. Costenbader KH, Feskanich D, Mandl LA, Karlson EW (2006) Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am J Med 119: 503.e501–509.
61. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, et al. (2003) Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis 62: 835–841.
62. Karlson EW, Mandl LA, Hankinson SE, Grodstein F (2004) Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheum 50: 3458–3467.
63. Guthrie KA, Dugowson CE, Voigt LF, Koepsell TD, Nelson JL (2010) Does pregnancy provide vaccine-like protection against rheumatoid arthritis? Arthritis Rheum 62: 1842–1848.
64. Spector TD, Hochberg MC (1990) The protective effect of the oral contraceptive pill on rheumatoid arthritis: an overview of the analytic epidemiological studies using meta-analysis. J Clin Epidemiol 43: 1221–1230.
65. Scott IC, Steer S, Lewis CM, Cope AP (2011) Precipitating and perpetuating factors of rheumatoid arthritis immunopathology: linking the triad of genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis. Best Pract Res Clin Rheumatol 25: 447–468.
66. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, et al. (2006) Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. Arthritis Res Ther 8: R133.

Figure S1. REGENT Stage 1-Simulation of General Population Risk Profiles: Example Using a Model Incorporating 3 SNPs

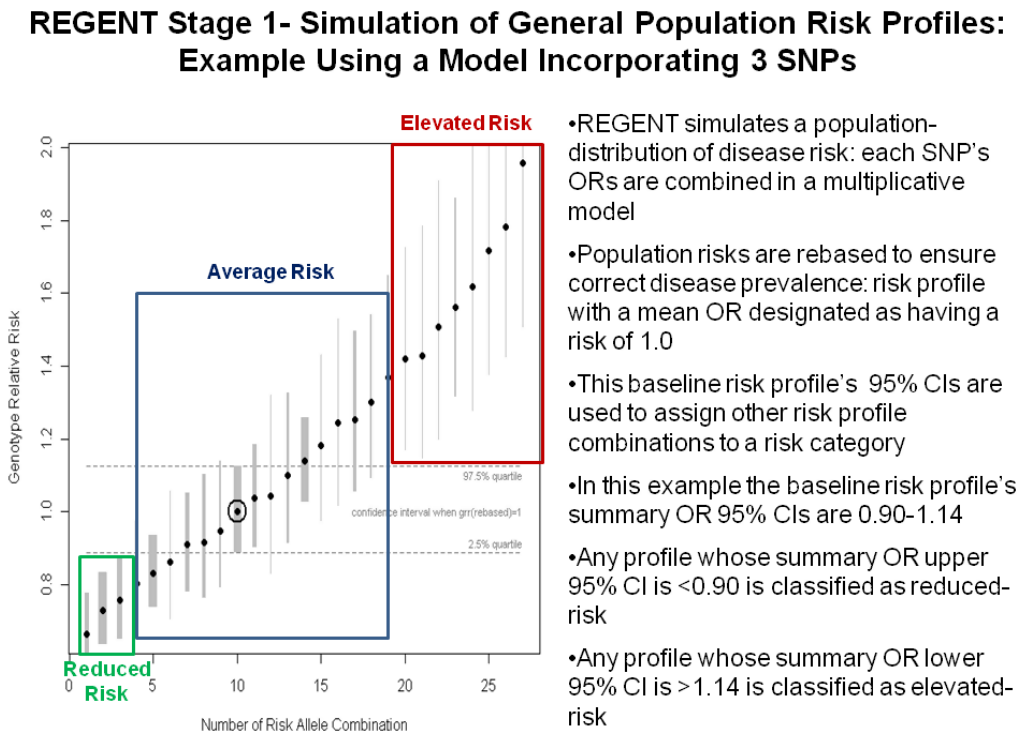


Figure S2. Main Findings of Each Prediction Model

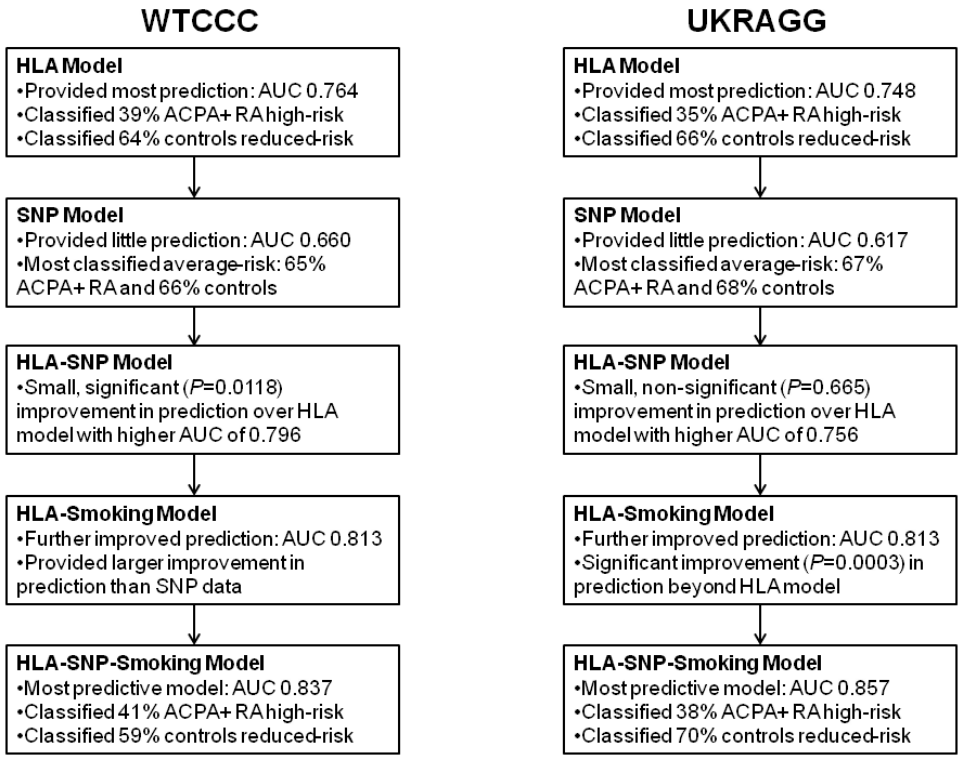


Table S1. Risk Categorisation Results for Models Incorporating Smoking in Females

Risk Category	**WTCCC HLA-Smoking Model**			**UKRAGG HLA-Smoking Model**			**WTCCC HLA-SNP-Smoking Model**			**UKRAGG HLA-SNP-Smoking Model**		
	Sero+ n=947	ACPA+ n=751	Controls n=736	Sero+ n= 1355	ACPA+ n=983	Controls n=606	Sero+ n=947	ACPA+ n=751	Controls n=736	Sero+ n= 333	ACPA+ n=182	Controls n=231
Reduced	268 (28.3)	179 (23.8)	437 (59.4)	417 (30.8)	264 (26.9)	387 (63.9)	202 (21.3)	127 (16.9)	401 (54.5)	84 (25.2)	38 (20.9)	137 (59.3)
Average	330 (34.8)	269 (35.8)	225 (30.6)	502 (37.0)	369 (37.5)	159 (26.2)	336 (35.5)	273 (36.4)	254 (34.5)	129 (38.7)	81 (44.5)	63 (27.3)
Elevated	19 (2.0)	17 (2.3)	6 (0.8)	35 (2.6)	25 (2.5)	10 (1.7)	107 (11.3)	88 (11.7)	32 (4.3)	30 (9.0)	16 (8.8)	14 (6.1)
High	330 (34.8)	286 (38.1)	68 (9.2)	401 (29.6)	325 (33.1)	50 (8.3)	302 (31.9)	263 (35.0)	49 (6.7)	90 (27.0)	47 (25.8)	17 (7.4)
AUC[a] *(95% CI)*	0.758 (0.733-0.782)			0.749 (0.724-0.773)			0.783 (0.760-0.806)			0.738 (0.690-0.786)		

Data are number (%) unless stated otherwise; [a] = AUCs calculated using ACPA-positive cases.

Table S2. Proxy SNPs Used In Modelling

WTCCC			
Meta-Analysis SNP (Stahl, et al., 2010)	**Loci**	**Proxy SNP Used**	**r^2**
rs10499194	*TNFAIP3*	rs13207033[a]	1
rs10488631	*IRF5*	rs12531711[a]	1
rs3761847	*TRAF1, C5*	rs10118357[a]	0.967
rs10865035	*AFF3*	rs9653442[b]	0.967
rs26232	*C5orf30*	rs35797[b]	0.901
rs3093023	*CCR6*	rs6907666[b]	0.933
rs706778	*IL2RA*	rs10795791[b]	0.934
rs5029937	*TNFAIP3*	rs5029939[b]	1
rs2476601	*PTPN22*	rs6679677[b]	1
rs13031237	*REL*	rs702873[c]	1
rs7574865	*STAT4*	rs3821236[c]	1
UKRAGG			
Meta-Analysis SNP (Stahl et al, 2010)	**Loci**	**Proxy SNP Used**	**r^2**
rs10499194	*TNFAIP3*	rs12527282[a]	1
rs3890745	*TNFRSF14*	rs10910099[a]	0.926

a = proxy SNP obtained using 1,000 Genomes CEU population panel (Johnson, et al., 2008); b = proxy SNP obtained using HapMap release 22 CEU population panel (Johnson et al, 2008); c = proxy SNP obtained using Ricopili (Broad Institute, Boston, USA) from the GWAS meta-analysis of RA risk (Ripke and Thomas, 2011).

Table S3. Two-Digit, Four-Digit and Mixed-Digit HLA Prediction Model Results

WTCCC									
	Two-Digit Resolution Model			**Four-Digit Resolution Model**			**Mixed-Digit Resolution Model**		
Risk Category	Sero+ n=1516	ACPA+ n=1061	Controls n=1647	Sero+ n=1342	ACPA+ n=966	Controls n=1126	Sero+ n=1516	ACPA+ n=1061	Controls n=1647
Reduced	440 (29.0)	267 (25.2)	973 (59.1)	386 (28.8)	242 (25.1)	678 (60.2)	471 (31.1)	283 (26.7)	1052 (63.9)
Average	108 (7.1)	78 (7.4)	140 (8.5)	135 (10.1)	102 (10.6)	132 (11.7)	301 (19.9)	218 (20.5)	304 (18.5)
Elevated	406 (26.8)	296 (27.9)	340 (20.6)	180 (13.4)	141 (14.6)	124 (11.0)	194 (12.8)	145 (13.7)	121 (7.3)
High	562 (37.1)	420 (39.6)	194 (11.8)	641 (47.8)	481 (49.8)	192 (17.1)	550 (36.3)	415 (39.1)	170 (10.3)
AUC (95% CI)	0.744 (0.726-0.763)			0.765 (0.745-0.785)			0.764 (0.746-0.782)		
UKRAGG									
	Two-Digit Resolution Model			**Four-Digit Resolution Model**			**Mixed-Digit Resolution Model**		
Risk Category	Sero+ n=2623	ACPA+ n=1508	Controls n=1500	Sero+ n=1534	ACPA+ n=1108	Controls n=735	Sero+ n=2623	ACPA+ n=1508	Controls n=1500
Reduced	802 (30.6)	409 (27.1)	931 (62.1)	342 (22.3)	231 (20.8)	393 (53.5)	844 (32.2)	430 (28.5)	987 (65.8)
Average	186 (7.1)	126 (8.4)	138 (9.2)	157 (10.2)	115 (10.4)	99 (13.5)	544 (20.7)	337 (22.3)	270 (18.0)
Elevated	732 (27.9)	409 (27.1)	288 (19.2)	200 (13.0)	151 (13.6)	82 (11.2)	388 (14.8)	215 (14.3)	118 (7.9)
High	903 (34.4)	564 (37.4)	143 (9.5)	835 (54.4)	611 (55.1)	161 (21.9)	847 (32.3)	526 (34.9)	125 (8.3)
AUC (95% CI)	0.743 (0.725-0.760)			0.743 (0.720-0.766)			0.748 (0.731-0.765)		

Data are number (%) unless stated otherwise; Sero+ = seropositive RA; ACPA+ = ACPA-positive RA; AUCs calculated using ACPA-positive cases; two-digit model evaluated individuals with all available HLA data collapsed down to two-digit resolution.

Table S4. Prediction Model Results: Risk Categorisation

WTCCC Prediction Models

Risk Category	HLA Model			SNP Model			HLA-SNP Model			HLA-Smoking Model			HLA-SNP-Smoking Model		
	Sero+ n=1516	ACPA+ n=1061	Controls n=1647	Sero+ n= 1516	ACPA+ n=1061	Controls n=1476	Sero+ n=1516	ACPA+ n=1061	Controls n=1476	Sero+ n= 287	ACPA+ n=239	Controls n=739	Sero+ n= 287	ACPA+ n=239	Controls n=739
Reduced	471 (31.1)	283 (26.7)	1052 (63.9)	167 (11.0)	109 (10.3)	355 (24.1)	329 (21.7)	186 (17.5)	880 (59.6)	57 (19.9)	36 (15.1)	442 (59.8)	47 (16.4)	30 (12.6)	434 (58.7)
Average	301 (19.9)	218 (20.5)	304 (18.5)	996 (65.7)	690 (65.0)	979 (66.3)	494 (32.6)	361 (34.0)	423 (28.7)	73 (25.4)	67 (28.0)	176 (23.8)	101 (35.2)	86 (36.0)	225 (30.4)
Elevated	194 (12.8)	145 (13.7)	121 (7.3)	226 (14.9)	165 (15.6)	102 (6.9)	167 (11.0)	108 (10.2)	75 (5.1)	59 (20.6)	47 (19.7)	64 (8.7)	30 (10.5)	25 (10.5)	30 (4.1)
High	550 (36.3)	415 (39.1)	170 (10.3)	127 (8.4)	97 (9.1)	40 (2.7)	526 (34.7)	406 (38.3)	98 (6.6)	98 (34.1)	89 (37.2)	57 (7.7)	109 (38.0)	98 (41.0)	50 (6.8)

UKRAGG Prediction Models

Risk Category	HLA Model			SNP Model			HLA-SNP Model			HLA-Smoking Model			HLA-SNP-Smoking Model		
	Sero+ n=2623	ACPA+ n=1508	Controls n=1500	Sero+ n=937	ACPA+ n=294	Controls n=573	Sero+ n=937	ACPA+ n=294	Controls n=573	Sero+ n=529	ACPA+ n=413	Controls n=322	Sero+ n=119	ACPA+ n=80	Controls n=115
Reduced	844 (32.2)	430 (28.5)	987 (65.8)	135 (14.4)	39 (13.3)	128 (22.3)	242 (25.8)	61 (20.7)	346 (60.4)	106 (20.0)	72 (17.4)	193 (59.9)	24 (20.2)	14 (17.5)	80 (69.6)
Average	544 (20.7)	337 (22.3)	270 (18.0)	633 (67.6)	198 (67.3)	390 (68.1)	319 (34.0)	112 (38.1)	150 (26.2)	127 (24.0)	96 (23.2)	90 (28.0)	33 (27.7)	24 (30.0)	27 (23.5)
Elevated	388 (14.8)	215 (14.3)	118 (7.9)	116 (12.4)	40 (13.6)	40 (7.0)	99 (10.6)	32 (10.9)	36 (6.3)	103 (19.5)	86 (20.8)	18 (5.6)	13 (10.9)	12 (15.0)	5 (4.3)
High	847 (32.3)	526 (34.9)	125 (8.3)	53 (5.7)	17 (5.8)	15 (2.6)	277 (29.6)	89 (30.3)	41 (7.2)	193 (36.5)	159 (38.5)	21 (6.5)	49 (41.2)	30 (37.5)	3 (2.6)

Data are number (%); Sero+ = seropositive RA; ACPA+ = ACPA-positive RA.

CHAPTER 7. DISCUSSION

7.1. Main Research Findings

This thesis has increased the knowledge of risk prediction in RA. This is an important step towards the implementation of a stratified approach to the management of this heterogeneous disease. The research in this thesis has enhanced the understanding of those factors that predict RA development, the subsequent clinical course of the disease, and the responses of patients to different treatment regimens. It has also demonstrated how these factors can be combined within a novel risk prediction modelling framework to estimate individuals' absolute risks of developing RA. This modelling could be used to stratify patients to groups that may benefit from preventive treatment strategies.

The findings from this thesis can be divided into five key areas. Firstly, it has improved the knowledge of which risk factors predict RA development by demonstrating that a significant inverse relationship exists between alcohol consumption and the risk of RA when the evidence across published studies is pooled using meta-analytical techniques. This suggests alcohol intake may protect against RA development. The research underlying this is outlined in Chapter 2 of this thesis.

Secondly, it has demonstrated an association between HLA susceptibility variants and radiological progression in a unique clinical trial cohort of early, active RA patients, whilst demonstrating that non-HLA susceptibility variants for RA and other immune-mediated diseases do not associate with this disease outcome in the same patient group. This suggests that the non-HLA genetic architectures of RA susceptibility and severity probably, at least in part, differ. The research underlying this is outlined in Chapters 3 and 4 of this thesis.

Thirdly, it has shown that patients with early, active ACPA-positive and ACPA-negative RA differ in their requirements for, and responses to, intensive combination DMARD and corticosteroid treatments. This suggests that ACPA status is an important biomarker for guiding treatment decisions in early, active RA patients. The research underlying this is outlined in Chapter 5 of this thesis.

Fourthly, it has demonstrated that estimating an asymptomatic individual's future risk of RA is possible, through developing and validating a risk prediction model that uses computer simulation to significantly improve upon the discriminative abilities of existing prediction models for this disease. This has provided a modelling framework that is applicable to many aspects of stratified medicine. The research underlying this is outlined in Chapter 6 of this thesis.

Finally, it has highlighted the importance of considering RA's heterogeneity when assessing its predictive factors, by demonstrating that a) the protective effect of alcohol on RA development is predominantly seen in ACPA-positive disease (Chapter 2); b) prediction modelling is more effective for ACPA-positive (as opposed to seropositive) disease (Chapter 6); and c) genetic and environmental factors have different impacts on the risk of developing younger and older onset RA, with genetic and environmental risk effects greater in YORA and elderly onset RA (EORA), respectively (Chapter 6).

7.2. Alcohol Consumption and RA Development

7.2.1. Principal Findings

Through systematically reviewing and performing a meta-analysis of published studies this thesis has demonstrated that a significant inverse relationship exists between alcohol consumption and the development of RA (Scott et al, 2013c). In those studies evaluating risk stratified by ACPA status a significant association was only observed for ACPA-positive RA. Taken together these findings indicate that alcohol intake probably protects against RA development, particularly ACPA-positive disease.

7.2.2. Strengths and Limitations

The strengths of this systematic review and meta-analysis are that it was performed in a highly standardised and systematic manner, testing a pre-specified hypothesis and according to a pre-defined protocol. Its main limitation was that causality could not be determined as a non-significant relationship was seen when the meta-analysis was restricted to cohort studies. There are many reasons why RA patients may

consume less alcohol, such as methotrexate use, and therefore asking patients to recall their past alcohol consumption – as would have been undertaken in most of the case-control studies – could be affected by recall bias; this could account for the inverse association observed. Another important limitation was that marked clinical and methodological heterogeneity existed between the included studies, which had significant variation in not only their findings but also how they defined and captured information on alcohol consumption. This limited the suitability of combining their results to provide an overall OR for RA.

7.2.3. Further Research

Since this literature review was performed a further two prospective cohort studies have been published (EPIC-2-NOAR and the NHS), which support its findings (Lahiri et al, 2014, Lu et al, 2014). Both studies reported risks using HRs. In EPIC-2-NOAR the age and sex adjusted HR for seropositive inflammatory arthritis for every 7 units of alcohol consumed/week was 0.83 (95% CI 0.69-0.98) (Lahiri et al, 2014). In the NHS the pooled multivariable adjusted HR for seropositive RA in alcohol drinkers consuming 5.0-9.9 grams/day compared to non-drinkers was 0.69 (95% CI 0.50-0.95) (Lu et al, 2014). As these studies used HRs to quantify risk they could not be included in a meta-analysis with the cohort studies identified by the published literature review. These two studies, however, provide further support for the concept that alcohol has a direct protective effect on RA development. As previously discussed this could be mediated via attenuation of the innate immune system (Jonsson et al, 2007, Mandrekar et al, 2006).

7.2.4. Clinical Implications

As most individual's risks of developing RA are low and alcohol consumption is associated with many adverse outcomes, it seems inappropriate to apply the finding of alcohol protecting against RA development in a public health context, in which asymptomatic individuals are advised to consume alcohol to prevent RA. The key clinical implication of this work is that data on alcohol consumption could be incorporated within risk prediction models to improve their ability to stratify asymptomatic individuals into risk groups for RA development. Identifying high-risk individuals would enable preventive strategies to be evaluated.

7.3. Genetic Markers for Radiological Progression

7.3.1. Principal Findings

This thesis has confirmed the influence of HLA susceptibility variants on radiological progression in RA, by demonstrating that the three main HLA risk variants (the *HLA-DRB1**04:01 allele and Val11 and His13 in the HLA-DRβ1 molecule) also associated with Larsen score progression in early, active RA patients in the CARDERA Genetics Cohort (CGC). This effect appeared to be independent of ACPA status, with similar effect sizes observed on restricting the analysis to ACPA-positive individuals. Conditional analyses indicated the association at both Val11 and His13 was driven by the tight LD between them, although it was not possible to establish which had the dominant association. Furthermore, conditional analyses also suggested their association was driven by a high correlation with the SE at positions 70-74. Due to the broad LD across the MHC region (De Bakker et al, 2006) alongside the modest size of the CGC it was not possible to further characterise which HLA amino acid positions had a SE-independent association with joint destruction.

This thesis has also demonstrated that no significant association exists between non-HLA susceptibility loci and radiological progression in the CGC, when evaluating these variants both individually and cumulatively, using a wGRS. Although the sample size of 524 patients is substantially smaller than the combined GWASs evaluated to identify these susceptibility loci (Okada et al, 2013), the CGC was adequately powered (at >80%) to detect genetic markers accounting for just 2% of the variance in X-ray progression. This indicates that non-HLA susceptibility loci do not have a clinically relevant association with radiological progression in early, active RA patients. It also suggests that the non-HLA genetic architectures of RA susceptibility and severity probably, at least in part, differ.

Finally, this thesis demonstrated no significant association between 138,488 genetic markers linked to a broad range of immune-mediated disorders present on the ImmunoChip and joint destruction rates in the CGC. When this analysis is considered alongside the two other similarly sized genome-wide studies, which have only identified 3 replicable associations with radiological progression in RA between

them (De Rooy et al, 2013b, Knevel et al, 2013b), it appears likely that the genetic architecture of radiological progression in RA is comprised of many variants of a small effect size. This suggests the optimal strategy to capture relevant variants is through the meta-analysis of GWASs.

7.3.2. Strengths and Limitations

This component of the thesis represents the first analysis of genetic associations with radiological progression in early, active RA patients within a clinical trial setting. As such it has a number of unique strengths. Firstly, as all patients had active disease it was able to identify genetic markers that associated with radiological progression independently of baseline disease activity. In contrast, existing studies use observational data from cohorts of patients with a range of disease activities; they are, therefore, unable to discern between genetic variants associating indirectly with radiological progression through influences on disease activity and those variants directly mediating radiological progression. Secondly, as the evidence base for using intensive combination therapy in RA is based on RCTs of patients with early, active disease (Boers et al, 1997a, Goekoop-Ruiterman et al, 2008, Mottonen, et al., 1999), if genetic predictors are to be of clinical value in guiding treatment decisions they too require evaluation in patients with early, active RA. Thirdly, all radiological assessments were performed in a timely, standardised manner. Fourthly, the regularity of Larsen scoring optimised the power to detect longitudinal associations. Fifthly, the randomisation to treatment groups ensured the effects of treatment could be fully adjusted for.

This analysis also has a number of important limitations. Foremost is its modest sample size of 524 patients, which limited the power to detect genetic associations of a small effect size. The analysis in Chapter 3, however, was well powered to detect variants accounting for $\geq$2% of the variance in radiological progression when applying a Bonferroni correction for testing 69 susceptibility variants; therefore, it was adequately powered to detect any susceptibility markers providing clinically useful information. The genome-wide study in Chapter 4 was substantially less well powered; it therefore probably contains a high false-negative rate. Secondly, only 24-month radiological data were available; genetic variants could exert their effects over

longer time periods. Thirdly, as with most longitudinal studies, outcome data were missing in some, albeit a minority, of patients. However, the linear mixed-effects model was able to accommodate missing data, by making the assumption that these were missing-at-random (MAR) (Higgins and Green, 2011). Additionally, similar results were obtained for the ANOVA model used in Chapter 3, which excluded individuals with missing Larsen score data. Fourthly, as the CGC comprised individuals from RCTs, radiological progression rates could be lower than those seen in observational studies, further reducing the analysis power. This seemed unlikely with similar proportions of CGC patients having clinically relevant Larsen score progression in comparison to observational studies of early RA patients (Sanmarti et al, 2007). Furthermore, the linear mixed-effects model was able to detect known associations with X-ray damage in the CGC, such as disease duration, ACPA status, and *04:01 carriage. Fifthly, although within the CGC radiological progression rates were linear, this is unlikely to persist over longer disease periods. Therefore, the β-values associated with each variant may only be applicable to the first few years of disease. Finally, several validated, non-HLA genetic associations with radiological progression identified in other studies were not replicated in the CGC. Although this further questions the power of the study, this non-replication is not unique. Within the Leiden ImmunoChip study only 14% of the identification cohort's significant SNPs present in the replication dataset attained statistical significance in this latter cohort (De Rooy et al, 2013b). Additionally, Knevel *et al* failed to replicate associations between *TRAF1-C5* and *TNFAIP3-OLIG3* variants and radiological progression across 6 independent cohorts (Knevel et al, 2012a). As discussed in Chapter 3, the two most likely explanations for this non-replication are firstly, that the CGC population − which comprises patients with early, severely active disease − is inherently different to the populations evaluated in these observational studies, which assessed patients with a range of disease durations and severities and secondly, the CGC looked at predictors of X-ray progression in the first 2 years of disease and other studies evaluated damage over longer time periods. Other explanations include the fact that previous studies were not able to fully adjust for the effects of treatment; some studies assessed genotype effects on the severity of X-ray damage, as opposed to its impact on progression over time; and the use of different radiological scoring systems across studies.

7.3.3. Further Research

There are four areas of further research, which are relevant to the candidate gene and genome-wide analyses undertaken in this thesis. Firstly, modern genotyped cohorts of RA patients with regular radiological scores present from diagnosis are needed. Nine observational study cohorts exist that have been widely used to identify genetic predictors of radiological progression in RA (Table 7-1). As most include patients with longstanding disease evaluated several decades ago their radiographic outcomes are likely to be substantially worse than those observed in currently diagnosed individuals and the relevance of their findings to contemporary practice is uncertain. Additionally, 4 of these studies only had a single X-ray score performed during established disease; they therefore often used estimated yearly progression (X-ray score divided by disease duration in years at the time of the radiograph) as the response variable in linear regression analyses. This makes the assumption that radiological progression is a linear trait, which whilst probably true in early disease (Hulsmans, et al., 2000) is unlikely to be valid in established RA. One observed pattern is that radiological progression is greatest in the first few years of disease and reduces thereafter (Lindqvist, et al., 2003) although a diverse range of progression patterns have been described (Plant, et al., 1998). This indicates a key research requirement to establish modern genotyped cohorts of RA patients, with radiological scores present at diagnosis and regular intervals thereafter. This will enable the most accurate modelling of X-ray progression and any positive associations are more likely to be of relevance to contemporary practice.

Table 7-1. Observational Studies Evaluating Genetic Associations with RA X-Ray Progression

Cohort	Size	Year Onset	Follow-Up	Outcome	X-ray Frequency	Main Statistical Methods	Loci Identified/Validated
Leiden EAC	600	1993–06	7 yrs	SvHS	Annual	Longitudinal linear model	*CD40, IL15, DKK1, IL2RA, GRZB, IL4, SPAG16, C5orf30, MMP9, ZFP36L1/C14orf181*
NARAC	385	1953–02	NA	SvHS	Single (duration 14 yrs)	Linear regression model	*CD40, IL4, SPAG16*
Groningen	280	1945–01	14 yrs	SvHS	Multiple (variable times)	Multivariate regression model	*IL15, DKK1, GRZB, IL4*
Lund	147	1985–90	5 yrs	Larsen	Annual	Multivariate regression model	*IL15, DKK1, GRZB, IL4*
Iceland	285	1942–08	NA	SvHS	Single (unspecified)	Linear regression model	*IL2RA*
Wichita	113	1963–99	10 yrs	SvHS	Multiple (variable times)	Multivariate regression model	*IL2RA, IL4, SPAG16, MMP9, ZFP36L1/C14orf181*
NDB	756	1980–99	NA	SvHS	Single (duration 12 yrs)	Linear regression model	*IL2RA, IL4, SPAG16, MMP9, ZFP36L1/C14orf181*
YEAR	418	2000-09	2 yrs	SvHS	Annual	ZINB evaluating baseline SvHS	*C5orf30*
GORA[a]	885	-	NA	Larsen	Single (duration 12 yrs)	ZINB/linear regression model	*C5orf30, IL15, DKK1, GRZB, IL4*

Loci and the relevant cohorts are described in the following papers: CD40 (Van Der Linden et al, 2009), IL15 (Knevel, et al., 2012a), *DKK1 (De Rooy, et al., 2013a), IL2RA (Knevel et al, 2013a), GRZB (Knevel, et al., 2013a), IL4 (Krabben et al, 2013), SPAG16 (Knevel et al, 2013b), C5orf30 (Teare et al, 2013), MMP9 (De Rooy et al, 2013b), ZFP36L1/C14orf181 (De Rooy et al, 2013b). EAC = Early Arthritis Clinic; NARAC = North American RA Consortium; NDB = National Data Bank; YEAR = Yorkshire Early Arthritis Register; GORA = Genetics of RA; Yrs = years; ZINB = zero-inflated negative binomial regression model; SvHS = Sharp-van der Heijde score; a = subset of 391-396 patients from the GORA cohort was used to evaluate IL15, DKK1, GRZB and IL4 loci.*

Secondly, meta-analyses of GWASs evaluating X-ray progression in RA are required to define its genetic architecture. The 12 validated genetic loci associations with radiological progression in RA have been estimated in the Leiden EAC to explain only 12-18% of the variance in joint destruction in this cohort (Van Steenbergen, et al., 2014); this indicates that a substantial proportion of the heritability of X-ray progression remains unexplained. As the genetic basis of radiological progression probably comprises multiple small effect size variants, the optimal strategy to define these is through the meta-analysis of GWASs. A collaborative study is required to undertake this in the existing available cohorts, although any findings will be limited by the heterogeneity across individual studies and the modelling assumptions required to undertaken a meta-analysis.

Thirdly, in view of the importance of the HLA region to joint destruction in RA an HLA-wide analysis of its association with X-ray progression across multiple large cohorts is needed, in a similar manner to that undertaken for RA susceptibility (Raychaudhuri et al, 2012) . The broad LD across the MHC alongside the fact that many amino acid polymorphisms external to positions 70-74 are correlated with the SE sequence indicates that very large sample sizes are required in order to dissect the relevant amino acid positions through conditional analyses.

Fourthly, further work is required to establish the heritability of other RA outcomes such as disease activity scores, disability levels and quality of life. The relevance of X-ray progression to contemporary practice is uncertain due to the lower levels of baseline radiographic damage observed in currently diagnosed RA patients (Rahman et al, 2011), the effect of combination DMARDs and biologics in limiting joint destruction (Goekoop-Ruiterman et al, 2008, Boers et al, 1997a, Graudal and Jurgens, 2010) and the limited correlation between radiological scores and other clinical responses (Strand and Sharp, 2003). If other RA outcomes are shown to have a significant genetic component then identifying their genetic predictors could deliver personalised care that is more relevant to current clinical practice and patients.

7.3.4. Clinical Implications

The findings from this thesis indicate that non-HLA genetic susceptibility variants for RA are unlikely to provide clinically informative data for guiding treatment decisions in early RA patients. Okada *et al* recently proposed that GWAS identified susceptibility loci could be harnessed to identify therapeutic agents currently used in other diseases, which could be repurposed for RA treatment (Okada et al, 2013). The findings from the CGC suggest this approach would be more successful in identifying preventative treatments, as opposed to drugs that modify the disease course of an established RA phenotype.

7.4. ACPA Status Predicts Treatment Requirements and Responses

7.4.1. Principal Findings

A secondary analysis of the CARDERA-1 trial demonstrated that only ACPA-positive, early, active RA patients benefited from intensive treatment with combination DMARDs and high-dose, tapering corticosteroids; no benefits beyond methotrexate monotherapy were observed in ACPA-negative patients. In this analysis intensive combination therapy was only needed to prevent radiological progression in ACPA-positive patients. Additionally, corticosteroids only provided significant improvements in disease activity and physical health outcomes in ACPA-positive RA. Taken together these findings suggest that current NICE guidelines advocating combination DMARDs and short-term corticosteroids in all patients presenting with active disease may result in the over-treatment of ACPA-negative individuals. They support a stratified approach to the management of early RA patients, based on prognostic markers like ACPA.

7.4.2. Strengths and Limitations

The strengths of this study included its large sample size (431 patients represents a reasonable size for an RCT), the involvement of multiple centers throughout the UK, which increases the generalisability of its findings, the measurement of a wide range of outcomes and the use of two-year follow-up data.

It has several limitations. As a secondary analysis of a published RCT it did not test a pre-specified primary hypothesis or use a pre-specified analytical plan. ACPA status was unknown in 8% of patients, who were excluded from the analysis. One of the study drugs, ciclosporin, is rarely used in current practice. Fewer ACPA-negative patients were studied; however, the power to detect a MCID in Larsen scores between combination therapy and monotherapy treatment arms in ACPA-negative patients was higher (86%) than in ACPA-positive patients (55%). Finally, the maximal dose of methotrexate was 15 mg/week; higher doses are often used in contemporary clinical care.

7.4.3. Further Research

Further work is required to evaluate the role of ACPA status in predicting firstly, the benefits of more commonly used combination DMARD regimes (such as methotrexate, hydroxychloroquine and sulfasalazine) and secondly, responses to low dose corticosteroids (such as intramuscular depomederone) in early, active RA patients. Additionally, there is a clear requirement for other reproducible prognostic and treatment response biomarkers to be identified because, as outlined in the introductory chapter of this thesis, current knowledge of these is limited.

In addition to the main analysis undertaken in this thesis, a subsidiary study was completed in the CARDERA-1 trial. This assessed if clinical and serological (IgM-RF and ACPA) markers predicted DAS28 defined remission rates in response to combination DMARDs. This work followed on from a previous systematic review by Ma *et al*, which demonstrated that combination DMARDs increased remission rates in early RA patients (Ma et al, 2010). The published results (Ma, et al., 2014) showed that gender, age, TJC, RF and ACPA status associated with an increased likelihood of attaining remission at 24-months. Patients that were male, aged over 50 years, had ≥6 tender joints, were RF-positive or ACPA-positive were more likely to achieve remission at 24-months when receiving triple therapy compared to monotherapy (ORs for remission of 2.99, 4.95, 2.71, 2.54 and 3.52, respectively). As this study represented a more restricted analysis of the CARDERA-1 outcome data at a single time point, and focused on remission only, it has not been incorporated within the main body of this thesis. However, it provides supportive data for the use

of serology and other clinical markers to predict intensive combination treatment responses.

7.4.4. Clinical Implications

The findings from this analysis, which are supported by data from the BeST study (De Vries-Bouwstra et al, 2008), suggest that ACPA is an important biomarker for stratifying early, active RA patients to treatment subgroups. It would be inappropriate to recommend that all ACPA-negative individuals receive DMARD monotherapy based on the results of the CARDERA-1 trial; although they show a lack of benefit beyond methotrexate monotherapy in ACPA-negative patients, it is likely that other combination treatments will have a positive impact in these individuals. However, if ACPA is considered alongside other poor prognostic markers, like the presence of radiological erosions, then the stratification of patients to treatment subgroups should be possible. This approach is already being undertaken in the USA, with current ACR guidelines recommending that combination DMARD use is restricted to patient subgroups with poor prognostic factors such as ACPA (Singh, et al., 2012). Although in England and Wales current NICE guidelines do not advocate such an approach, a recent national audit of UK rheumatologists reported that only 50% used initial combination therapy in patients with newly diagnosed RA; poor prognostic markers like erosions and ACPA positivity were key factors in determining whether combination therapy was used (Garrood et al, 2011). The clinical implications of this study are, therefore, that stratified medicine in RA is possible, with ACPA status representing one such method of stratifying patients into subgroups that are more, or less likely to benefit from combination treatments.

7.5. Improved Risk Prediction Modelling for RA

7.5.1. Principal Findings

The study outlined in Chapter 6 of this thesis demonstrated that it is possible to use validated RA risk factors to estimate an individual's lifetime risk of developing seropositive RA. This is an important first step in evaluating primary preventive strategies in high-risk individuals. The simulated approach used in this analysis

provided conceptual advances in risk prediction modelling to improve upon the discrimination of existing prediction models. The highest AUC for a RA gene-environment prediction model developed in this thesis comprised 0.86 (Scott et al, 2013b); the highest previously published AUC for a RA gene-environment prediction model is 0.80 (Yarwood et al, 2013).

7.5.2. Strengths and Limitations

There were two main strengths to this study. Firstly, the prediction models were validated in two large, independent datasets (WTCCC and UKRAGG) comprising approximately 4,000 cases and 3,000 controls. Highly similar results were observed in both cohorts. Secondly, as only validated susceptibility factors were incorporated in the prediction models, with summary ORs obtained from large meta-analyses, their generalisability across patient populations was optimised. This study also had a number of limitations. Firstly, WTCCC participants were included in the meta-analyses that genetic risk loci data were obtained from; however WTCCC comprised only a proportion of the meta-analyses datasets and the findings were independently replicated in UKRAGG. Secondly, missing data meant the number of individuals evaluated by each model fell as more risk factors were included. Thirdly, due to marked heterogeneity in published data on epistasis/gene-environment interactions independence between factors was assumed; this oversimplification fails to consider important interactions such as those existing between the SE alleles and smoking.

There are also several important limitations in all risk prediction models that aim to estimate asymptomatic individuals' risks of RA. Firstly, they use ORs, which are derived from cross-sectional case-control studies, to approximate risk. ORs and RRs are not, however, interchangeable; if an OR is considered in the same manner as a RR it will always overestimate the effect size (Davies, et al., 1998). Secondly, they are only able to provide information on the lifetime odds or absolute risks of disease, as opposed to data on the risk of developing RA over a specific time period. A more clinically useful model would provide information on the absolute risk of RA over the next few years, in a manner akin to the predictive data generated by the Framingham Heart Study calculator for future risk of cardiovascular events (D'agostino, et al., 2008). This would allow the evaluation of preventive measures

over a more realistic time period than a patient's entire lifetime. Unfortunately, such modelling is not currently possible as no prospective cohort study exists that has genetic data on the entire dataset to derive time specific risks. Thirdly, in the general population the absolute risks of developing RA are low and it is only a small minority of individuals that are at a high-risk of disease. This is to be expected under Bayes' Theorem, in which the post-test odds of developing a disease are not just influenced by the likelihood ratio (probability of test result in a diseased person/probability of test result in a non-diseased person) of a test but also the pre-test odds of developing the disease (Prince, 1996). In the context of RA the pre-test probability can be considered the same as the disease prevalence, which is low at approximately 1%. Therefore, regardless of the prediction model's likelihood ratio, the post-test probability of developing RA will be small in most individuals owing to its low prevalence. It is unlikely that identifying further risk factors to incorporate within prediction models will substantially alter this and RA prediction models are, therefore, unsuitable for use as screening tools within the general population.

7.5.3. Further Research

There are three key areas of further research arising from this study. Firstly, the devised prediction models require evaluating in *a priori* high-risk populations. In this context they may identify clinically relevant proportions of high-risk people enabling preventive measures to be evaluated. One high-risk population is first-degree relatives (FDRs) of RA patients. Recently published data from Sweden suggests the familial OR for RA is approximately 3 in first-degree relatives (Frisell, et al., 2013); this may increase if several family members are affected. Such work is already underway, with the PRe-clinical EValuation of Novel Targets in RA (PREVeNT RA) study, led by Professor Ian Bruce (University of Manchester), currently establishing a UK based cohort of FDRs of RA patients (United Kingdom Clinical Research Network, 2013). The aim of this study is to use gene-environment prediction models to stratify FDRs by RA risk profiles and to undertake serological and proteomic profiling alongside ultrasound imaging in high-risk individuals to evaluate the evolution of RA. An overview of this study is provided in Table 7-2.

Table 7-2. Overview of the PREVeNT RA study

Aims
1. Establish a national cohort of FDRs of RA patients 2. Use clinical-genetic risk prediction models to stratify FDRs according to disease risk 3. Use serological/proteomic profiling and ultrasound imaging to evaluate disease evolution in high-risk FDRs 4. Identify incident cases of IA 5. Study cardiovascular risk biomarkers in a subset of this cohort.
Rationale
RA is the commonest chronic IA with high costs to the UK economy. Gene-environment risk factors interact to influence RA development. The identification of high-risk individuals in the community could facilitate preventive treatments.
Subjects
Individuals will be included that are aged ≥30 years, UK residents, willing to provide consent and complete questionnaires, willing to provide a blood sample, willing to inform the study centre if they develop symptoms of RA, are an FDR of a proband with a diagnosis of RA from a Rheumatologist.
Sample Size
The recruitment target is 3,000 individuals.

FDR= first degree relative, IA = inflammatory arthritis; Information on the PREVeNT RA study obtained from the UK Clinical Research Network Study Portfolio (United Kingdom Clinical Research Network, 2013).

Secondly, the optimal way in which predictive data is conveyed to patients requires determining. A Cochrane review reported little or no effect of communicating genetic-based risk estimates on lifestyle modifications to reduce disease risk (Marteau, et al., 2010). The quality of the 14 included studies was, however, judged to be weak; additionally they focused on communicating DNA-based risk information in the form of the presence or absence of a specific risk variant, as opposed to communicating the absolute risks of disease development. The effects of communicating lifetime disease risks (as provided by the models described in this thesis), alongside the manner in which these are conveyed (for example face-to-face or telephone delivery) requires evaluation.

Thirdly, the possible preventive strategies that could be used needs careful consideration. Simple lifestyle modifications, such as smoking cessation, are likely to substantially reduce the risk of ACPA-positive disease, particularly in SE carriers. The ethical considerations of studies assessing the role of immunomodulatory

treatments, such as biologic drugs, to prevent RA in entirely asymptomatic high-risk individuals are, however, considerable. Data on possible non-conservative strategies for primary prevention will be generated by studies evaluating secondary prevention treatments. One example of this is the Arthritis Prevention In the Pre-clinical Phase of RA with Abatacept (APIPPRA) study, led by Professor Andrew Cope (King's College London). This double blind, placebo controlled clinical trial will evaluate the role of abatacept therapy in preventing RA in symptomatic subjects (arthralgia without joint swelling) at high-risk of developing RA (e.g. those with ACPA present). This study will test the hypothesis that weekly subcutaneous abatacept injections over 12 months will substantially reduce the proportion of individuals developing a clinically apparent inflammatory arthritis.

7.5.4. Clinical Implications

The high individual and societal costs of RA mean that preventing its development is an important research goal. Although it is difficult to capture the precise costs of RA, an analysis undertaken in 2001 in America reported an annual cost per RA patient of 11,341 US dollars (Lajas et al, 2003). In the biologics era this cost will be substantially higher. Within the UK, in 2009 the National Audit Office estimated the annual cost of RA to the NHS to be approximately 560 pounds sterling (National Audit Office, 2009). The modelling framework developed in this thesis is an important first step towards the primary prevention of RA. Targeted screening of *a priori* high-risk groups like FDRs of RA patients could result in enough high-risk individuals being identified to enable the assessment of primary prevention strategies like smoking cessation. It would also enable a better understanding of the pre-clinical changes that lead to RA, through assessments such as the serological and proteomic profiling planned in the PREVeNT RA Study.

Additionally, the prediction modelling framework used in this thesis can easily be applied to a variety of other branches of stratified medicine. Once the predictive factors underlying radiological progression and treatment responses in RA are identified, this modelling approach could be used to generate a summary OR and risk category for probable X-ray progression or medication responses, which could be used to inform clinical decision-making.

7.6. The Importance of RA Heterogeneity

7.6.1. ACPA-Positive versus ACPA-Negative Disease

This thesis has highlighted the importance of ACPA status in determining RA subsets by demonstrating that firstly, the inverse association between alcohol consumption and RA risk is only significant for ACPA-positive disease; secondly, ACPA-positive and ACPA-negative RA have different requirements for intensive combination treatments; and thirdly, prediction modelling has better discrimination for ACPA-positive (as opposed to seropositive) RA. This adds to the accumulating evidence that ACPA-positive and ACPA-negative RA probably represent distinct disease entities, grouped together under the umbrella term of "RA". It provides a reasonable explanation for some of the contrasting results seen across studies examining predictive factors in RA, which often examine RA as a single group, and highlights a key research requirement to evaluate these subsets separately. This process is already beginning as any studies using the 2010 ACR RA classification criteria will favour the inclusion of ACPA-positive patients. It is, however, important that ACPA-negative RA is not overlooked as despite being a more difficult subset to define, which is at a higher risk of misclassification, it comprises a significant proportion of patients seen in rheumatology clinics.

7.6.2. Younger versus Older Onset RA

Traditionally YORA and EORA are defined categorically using an age cut-off for disease onset of ≤60 years and >60 years, respectively (Deal, et al., 1985). There is some evidence that YORA and EORA differ phenotypically, with three studies reporting EORA to have a more abrupt onset, characterised by higher rates of large joint involvement (Deal et al, 1985, Van Der Heijde, et al., 1991, Bajocchi, et al., 2000). They may also differ with respect to their disease course and treatment responses. An analysis of NOAR demonstrated a linear relationship between a reducing age of onset and remission, although when age of onset was dichotomised no association was seen (Harrison, et al., 2000). Similarly, an analysis of the Dutch biologics registry reported that anti-TNF was less effective in older RA patients (a substantial proportion of whom are likely to have EORA) (Radovits, et al., 2009).

Using an arbitrary age cut-off of 60 years to define YORA and EORA could result in the loss of important information on the extreme age of onset phenotypes. To overcome this, within the WTCCC and UKRAGG analysis outlined in this thesis, age of onset was evaluated as a continuous variable (Scott et al, 2013b). The resultant finding that high HLA-derived genetic risk scores associated with a younger age of RA onset and ever-smoking associated with an older age of RA onset, suggested that genetic and environmental factors are more important for the development of YORA and EORA, respectively. This novel finding has not been previously reported. As it was demonstrated in two large, independent datasets it probably represents a true finding; furthermore, as previously discussed it has biological plausibility. These findings are also supported by a recent register-based nested case-control study in Sweden, which evaluated the familial aggregation of RA across 3 large, independent datasets (Frisell et al, 2013). A substantially higher familial OR for RA was observed in YORA (defined as an age of onset <40 years). This was particularly true in seropositive disease: in the Swedish Rheumatology Quality Register the familial OR (risk of disease if an FDR is affected) for RF-positive RA with an age of onset <40 years and >60 years comprised 6.0 (95% CI 4.3-8.4) and 2.8 (95% CI 2.2-3.7), respectively; in the EIRA cohort the familial OR for ACPA-positive RA with an age of onset <40 years and >60 years comprised 6.2 (95% CI 3.5-10.9) and 3.3 (95% CI 2.0-5.3), respectively (Frisell et al, 2013). As familial risk is often considered to be mainly due to genetic factors, these findings suggest that genetic risks have a larger impact on YORA. Overall, the findings from this thesis add weight to the concept that age of RA onset may be an important way in which RA patients can be sub-classified.

7.7. Future Research

There are two areas of further research arising directly from the work undertaken in this thesis. Firstly, in order to increase the power to detect genetic associations with radiological progression in RA, a collaborative network has been established with researchers in the Rheumatic Disease Epidemiology Group (Dr Jing Cui and Dr Elizabeth Karlson) at Harvard Medical School (Boston, USA). Secondly, in order to define gene-environment RA risk factors in non-European individuals, a unique

observational study has been developed, entitled the GENetics of Ra in individuals of African ancestry (GENRA) study.

7.7.1. Identifying Genetic Predictors of Radiological Progression

The aim of this study is to undertake a genome-wide analysis evaluating genetic associations with the progression of radiological scores in the CGC and Brigham and Women's Rheumatoid Arthritis Sequential Study (BRASS) (Cui, et al., 2014). The proposal for this study is outlined in Table 7-3.

Table 7-3. GWAS Meta-Analysis of Genetic Predictors of X-Ray Progression

Aim
To identify genetic associations with X-ray progression in RA patients using a genome-wide approach across two independent datasets (CGC and BRASS).
Rationale
Candidate-gene approaches to identify genetic associations with radiological progression have important limitations. Genome-wide analyses in large datasets are most likely to identify replicable genetic associations.
Subjects
CGC: comprises 524 patients with early, active RA previously enrolled to two clinical trials. Modified Larsen scores are available at 6 to 12-month time points. *BRASS:* comprises 422 ACPA-positive established RA patients enrolled to a prospective observational study. SvHS are available on all patients at study enrolment (mean disease duration 17 years).
Genotyping
CGC was genotyped on the ImmunoChip. BRASS was genotyped on the Affymetrix 6.0 platform. Both datasets have been imputed to genome-wide coverage using IMPUTE2 (Howie, et al., 2009), based on 1,000 Genomes Phase 1 haplotypes.
Statistical Analysis
The CGC will be evaluated using the same strategy as outlined in Chapter 4 of this thesis. BRASS will be analysed as follows: estimated annual X-ray progression rates will be calculated by dividing the total SvHS by disease duration (in years) at the time of the radiograph. The association between log-transformed estimated SvHS annual progression rates and genotype will be tested in a linear regression model. Adjustments for other relevant variables will be made based on their association with X-ray progression in this dataset. *P*-values from the CGC and BRASS will be combined using meta-analysis techniques to provide an overall *P*-value for radiological progression for each SNP.
Replication Cohort
Any associations attaining P_{GWAS} will require replication. Suitable replication cohorts are in the process of being established.

7.7.2. Evaluating RA Predictive Factors in Non-European Populations

Studies of predictive factors in RA have mainly been undertaken in European or North American populations, in which most individuals are of European ancestry. This is exemplified in the meta-analysis of smoking as an RA risk factor; all of the 16 included studies were performed in Europe or the USA (Sugiyama et al, 2010). The application of these predictive factors to other ethnic groups is uncertain.

Within South London a significant proportion of patients seen in rheumatology clinics are of African ancestry; many of these individuals were either born in Africa or the Caribbean, with the latter patient group often having genetic admixture. Very little information on predictive factors in African ancestry RA patients exists. Although RA is predominantly a European disease, it does represent a significant healthcare problem in Africa, particularly in urban regions. A recent cross-sectional study in Kinshasa – an urban area in the Democratic Republic of Congo – reported the prevalence rate of RA (fulfilling the 1987 ACR classification criteria) was 0.6-0.9%, which is only slightly below that reported in European populations (Malemba, et al., 2012).

The most comprehensive assessment of risk factors for RA in African ancestry individuals has been undertaken in African Americans recruited to the Consortium for the Longitudinal Evaluation of African Americans with Early RA (CLEAR) registry. In this dataset, cigarette smoking was a significant RA risk factor; the age and sex adjusted OR for RA in ever-smokers was 1.45 (95% CI 1.07-1.97) (Mikuls, et al., 2010). The CLEAR registry has also demonstrated that the SE is an important RA risk factor in African Americans, providing an OR for RA of 2.35 (95% CI 1.84–3.00) (Hughes, et al., 2008). Its prevalence of 25.2% was, however, substantially lower than the 50-70% often reported in European RA patients, indicating that non-SE factors may be more important for precipitating disease in African ancestry individuals (Hughes et al, 2008). Furthermore, the presence of SE alleles was associated with a higher degree of European ancestry (assessed using a panel of >1,200 ancestry-informative markers); for every 1% increase in European ancestry the odds of carrying an *HLA-DRB1**04:01 allele increased by a factor of 1.035 (95% CI 1.004-1.068) (Hughes et al, 2008). It therefore appears likely that risk factor

similarities between European and African American individuals are, at least in part, driven by European admixture in the latter patient group.

Research evaluating RA risk factors in African ancestry individuals born in Africa is limited to a few small observational studies. In one case-control study of 56 Cameroonian RA patients and 50 healthy controls (medical students and hospital workers) attending an outpatient unit in Yaoundé (Cameroon) the SE was significantly commoner in cases (30% SE-positive) compared with controls (10% SE-positive) (Singwe-Ngandeu, et al., 2010). Its prevalence was, however, substantially lower than in Europeans and no individual carried *04:01, which is the commonest SE allele in Europeans (Raychaudhuri et al, 2012). As seropositivity rates were similar to European RA patients these findings suggest that the SE alleles play a lesser role in ACPA formation in Cameroonian RA patients. Another small study of 34 Senegalese RA patients and 220 controls reported an increased risk of RA in *HLA-DR3* and *HLA-DR10* carriers but not *HLA-DR4* carriers (Dieye, et al., 1997). Regarding non-HLA genetic associations, several small studies have used a candidate gene approach to test non-HLA susceptibility loci in sub-Saharan African (Tunisian (Chabchoub, et al., 2009, Ben Hamad, et al., 2012, Chabchoub, et al., 2006) and Egyptian (Mohamed, et al., 2012)) and Black South African (Moodley, et al., 2010) populations; their findings are generally negative. The most comprehensive assessment of non-HLA loci was carried out in 44 Cameroonian cases and 163 West/Central Africa controls. No association was observed between a wGRS formed from 28 validated European non-HLA susceptibility SNPs and RA in Africans (OR 0.71; 95% CI 0.29-1.74; P=0.456) (Viatte, et al., 2012).

There is, therefore, an important research need to better define environmental and genetic RA risk factors in African ancestry populations. As part of the Clinical Research Fellowship underlying this thesis an observational cross-sectional study has been developed, entitled the GENetics of Ra in individuals of African ancestry (GENRA) study. The primary aim of this study is to evaluate whether environmental and genetic risk factors for RA identified in European ancestry populations also associate with RA in African ancestry individuals living in the UK. Its details are outlined in Table 7-4. To date 194 patients have been recruited.

Table 7-4. Genetics of RA in Individuals of African Ancestry (GENRA) Study

Aim
To evaluate whether environmental and genetic risk factors for RA identified in European ancestry individuals also represent risk factors in African ancestry individuals living in the UK.
Rationale
Most studies of predictive factors in RA have focussed on European and North American populations. Little is known about how these risk factors apply to African ancestry individuals.
Subjects
This study will enrol patients of African ancestry from rheumatology outpatient clinics in four South London rheumatology units. *Inclusion Criteria:* a) A clinical diagnosis of RA by the 1987 and/or 2010 American College of Rheumatology RA classification criteria b) Self-reported ethnicity that is "Black/African/Caribbean/Black British" (classified according to the 2011 Census ethnicity grouping). *Exclusion Criteria:* a) Diagnoses of arthropathies other than RA b) Self-reported ethnicity other than "Black/African/Caribbean/Black British". Environmental risk factor data (on smoking, alcohol and BMI) has been obtained from 859 Black African and 1,067 Black Caribbean controls living in the UK that contributed to the Health Survey for England (HSE) 2004 ethnic minorities boost sample. Genotype data will be used from European ancestry RA patients within the WTCCC and CGC datasets.
Outcome Assessments
Information is captured on: a) Demographics- age, gender, ethnic grouping, ethnic grouping of parents/grandparents b) Disease duration c) Disease activity- DAS28 scores d) Radiological erosions e) Serology and inflammatory factors- the presence of autoantibodies (such as RF and ACPA) and other inflammatory factors (such as the ESR) f) Extra-articular features g) Disability- measured using HAQ h) Quality of Life- measured using EuroQoL i) Drug treatments- current and previous medications j) Co-morbidities k) Clinical risk factors for RA- smoking status, alcohol consumption and BMI data is captured at the time of assessment, time of RA diagnosis and 10 years before RA diagnosis through patient interview
Genotyping
This will be undertaken on the Illumina HumanOmniExpress BeadChip, which genotypes 733,202 markers. This platform provides adequate coverage of common genome-wide variation in African ancestry individuals; for common

SNPs (MAF>5%) it has coverage with an r^2 >0.8 of 73% in the 1,000 genomes African population panel (Nelson, et al., 2013).
Statistical Analysis
Environmental Risk Factors Adjusted ORs for RA in GENRA cases compared with HSE controls will be calculated with regards to smoking status, alcohol consumption and BMI. The first two variables are established RA risk factors in European populations; a link with RA in European studies has been reported in several studies for the latter factor. Initially cases and controls will be evaluated irrespective of their ethnicity; subsequent analyses will be undertaken stratified by ethnic grouping. *Genetic Risk Factors* A wGRS will be calculated for a) GENRA cases, b) European WTCCC cases, c) European CGC cases using validated European susceptibility loci identified by Okada *et al* (Okada et al, 2013) in the same manner as described in chapter 3 of this thesis. The cumulative distributions of the wGRSs will be compared across these populations to establish the degree of overlap in genetic risks between African and European ancestry RA patients.

BMI = body mass index; WTCCC = Welcome Trust Case Control Consortium; CGC = CARDERA Genetics Cohort; wGRS = weighted genetic risk score.

7.8. Conclusion

This thesis has demonstrated that risk prediction in RA is possible. It has advanced the knowledge of which factors predict RA development, the subsequent clinical course of the disease, and the responses of patients to different treatment regimens. It has shown how these predictive factors can be combined within a novel risk prediction modelling framework to stratify individuals to disease risk groups that may benefit from preventive treatments. The research in this thesis further highlights the requirement to adopt a stratified approach to RA management, using clinical characteristics and biomarkers such as ACPA to identify groups of individuals that are most likely to respond to specific treatment strategies.

There are several clinical implications from the research presented in this thesis. Firstly, the analysis of the CARDERA-1 trial data suggests that not all early, active RA patients require identical treatment. The results provide a strong case for focusing the use of intensive combination therapy in ACPA-positive disease and support using DMARD monotherapy in at least a proportion of ACPA-negative patients. Secondly, the analysis of the CGC indicates that non-HLA RA susceptibility variants do not provide clinically useful prognostic information in

early, active RA patients, with no association observed between these variants and radiological progression over 2 years. This suggests that GWAS derived susceptibility loci may be more useful in identifying treatment pathways relevant to disease prevention as opposed to the treatment of an established RA phenotype. Thirdly, an analysis of risk prediction modelling in the WTCCC and UKRAGG cohorts has demonstrated that identifying individuals at a high-risk of developing RA is possible; this raises the possibility that primary prevention strategies may be evaluated in the future.

There are also a number of research implications arising from this work. Firstly, it has confirmed the importance of considering disease heterogeneity when undertaking research in RA. RA subsets defined by both ACPA status and age of onset differ in their susceptibility factors; a failure to evaluate these subsets separately could explain the non-replication of many predictive factors across RA patient cohorts. Additionally, the fact that ACPA status determines treatment requirements and responses raises the question of whether clinical trials of RA treatments need to assess these subsets separately. Secondly, the genome-wide analysis of radiological progression predictors suggests that the genetic architecture of X-ray progression in RA comprises many variants of a small effect size. This highlights a research requirement to undertake GWAS meta-analyses in order to optimise the power to detect relevant associations. A collaborative network has been established with researchers at Harvard Medical School to undertake such an analysis in the CGC and BRASS datasets. Thirdly, as RA risk prediction models are unsuitable for population level screening (due to the low numbers of high-risk individuals that would be identified) their impact in *a priori* high-risk groups like FDRs of RA patients requires evaluation. In this context they could identify large enough numbers of high-risk individuals to allow primary prevention strategies to be evaluated.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., Depristo, M. A., Durbin, R. M., et al., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491, pp.56-65.

Abou-Raya, S., Abou-Raya, A., Naim, A. & Abuelkheir, H., 2008. Rheumatoid arthritis, periodontal disease and coronary artery disease. *Clin Rheumatol,* 27, pp.421-7.

Abraham, G., Tye-Din, J. A., Bhalala, O. G., Kowalczyk, A., Zobel, J. & Inouye, M., 2014. Accurate and robust genomic prediction of celiac disease using statistical learning. *PLoS Genet,* 10, pp.e1004137.

Adab, P., Jiang, C. Q., Rankin, E., Tsang, Y. W., Lam, T. H., Barlow, J., et al., 2014. Breastfeeding practice, oral contraceptive use and risk of rheumatoid arthritis among Chinese women: the Guangzhou Biobank Cohort Study. *Rheumatology (Oxford),* [Epub Ahead of Print].

Ahlmen, M., Svensson, B., Albertsson, K., Forslind, K., Hafstrom, I. & Barfot Study Group, 2010. Influence of gender on assessments of disease activity and function in early rheumatoid arthritis in relation to radiographic joint damage. *Ann Rheum Dis,* 69, pp.230-3.

Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., Bingham, C. O., 3rd, et al., 2010. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum,* 62, pp.2569-81.

Andermann, A., Blancquaert, I., Beauchamp, S. & Dery, V., 2008. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ,* 86, pp.317-9.

Anderson, J. J., Wells, G., Verhoeven, A. C. & Felson, D. T., 2000. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. *Arthritis Rheum,* 43, pp.22-9.

Andersson, A. K., Li, C. & Brennan, F. M., 2008. Recent developments in the immunobiology of rheumatoid arthritis. *Arthritis Res Ther,* 10, pp.204.

Arkema, E. V., Karlson, E. W. & Costenbader, K. H., 2010. A prospective study of periodontal disease and risk of rheumatoid arthritis. *J Rheumatol,* 37, pp.1800-4.

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., et al., 1988. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum,* 31, pp.315-24.

Atreya, I., Schimanski, C. C., Becker, C., Wirtz, S., Dornhoff, H., Schnurer, E., et al., 2007. The T-box transcription factor eomesodermin controls CD8 T cell activity and lymph node metastasis in human colorectal cancer. *Gut,* 56, pp.1572-8.

Bajocchi, G., La Corte, R., Locaputo, A., Govoni, M. & Trotta, F., 2000. Elderly onset rheumatoid arthritis: clinical aspects. *Clin Exp Rheumatol,* 18, pp.S49-50.

Banal, F., Dougados, M., Combescure, C. & Gossec, L., 2009. Sensitivity and specificity of the American College of Rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis. *Ann Rheum Dis,* 68, pp.1184-91.

Barnetche, T., Constantin, A., Cantagrel, A., Cambon-Thomsen, A. & Gourraud, P.-A., 2008. New classification of HLA-DRB1 alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. *Arthritis Res Ther,* 10, pp.R26.

Bartok, B. & Firestein, G. S., 2010. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol Rev,* 233, pp.233-55.

Beasley, T. M., Erickson, S. & Allison, D. B., 2009. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav Genet,* 39, pp.580-95.

Ben Hamad, M., Cornelis, F., Maalej, A. & Petit-Teixeira, E., 2012. A Tunisian case-control association study of a 6q polymorphism in rheumatoid arthritis. *Rheumatol Int,* 32, pp.1849-50.

Bengtsson, C., Nordmark, B., Klareskog, L., Lundberg, I., Alfredsson, L. & Group, E. S., 2005. Socioeconomic status and the risk of developing rheumatoid arthritis: results from the Swedish EIRA study. *Ann Rheum Dis,* 64, pp.1588-94.

Berglin, E., Johansson, T., Sundin, U., Jidell, E., Wadell, G., Hallmans, G., et al., 2006. Radiological outcome in rheumatoid arthritis is predicted by presence of antibodies against cyclic citrullinated peptide before and at disease onset, and by IgA-RF at disease onset. *Ann Rheum Dis,* 65, pp.453-8.

Berglin, E., Kokkonen, H., Einarsdottir, E., Agren, A. & Rantapaa Dahlqvist, S., 2010. Influence of female hormonal factors, in relation to autoantibodies and genetic markers, on the development of rheumatoid arthritis in northern Sweden: a case-control study. *Scand J Rheumatol,* 39, pp.454-60.

Bland, J. M. & Altman, D. G., 2000. Statistics notes. The odds ratio. *BMJ,* 320, pp.1468.

Boers, M., Verhoeven, A. C., Markusse, H. M., Van De Laar, M. A., Westhovens, R., Van Denderen, J. C., et al., 1997a. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet,* 350, pp.309-18.

Boers, M., Verhoeven, A. C., Markusse, H. M., Van De Laar, M. A., Westhovens, R., Van Denderen, J. C., et al., 1997b. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet,* 350, pp.309-18.

Boini, S. & Guillemin, F., 2001. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. *Ann Rheum Dis,* 60, pp.817-827.

Bos, W. H., Dijkmans, B. a. C., Boers, M., Van De Stadt, R. J. & Van Schaardenburg, D., 2010a. Effect of dexamethasone on autoantibody levels and arthritis development in patients with arthralgia: a randomised trial. *Ann Rheum Dis,* 69, pp.571-4.

Bos, W. H., Wolbink, G. J., Boers, M., Tijhuis, G. J., De Vries, N., Van Der Horst-Bruinsma, I. E., et al., 2010b. Arthritis development in patients with arthralgia is

strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. *Ann Rheum Dis,* 69, pp.490-4.

Breedveld, F., 2011. The value of early intervention in RA--a window of opportunity. *Clin Rheumatol,* 30 Suppl 1, pp.S33-9.

Brennan, P., Harrison, B., Barrett, E., Chakravarty, K., Scott, D., Silman, A., et al., 1996. A simple algorithm to predict the development of radiological erosions in patients with early rheumatoid arthritis: prospective cohort study. *BMJ,* 313, pp.471-6.

Brennan, P. & Silman, A. J., 1994. An investigation of gene-environment interaction in the etiology of rheumatoid arthritis. *Am J Epidemiol,* 140, pp.453-60.

Brik, R., Lorber, M., Rivkin, M. & Nahir, A. M., 1990. ELISA determined IgM and IgA rheumatoid factors in seronegative rheumatoid and psoriatic arthritis. *Clin Exp Rheumatol,* 8, pp.293-6.

Brown, W. M., Pierce, J., Hilner, J. E., Perdue, L. H., Lohman, K., Li, L., et al., 2009. Overview of the MHC fine mapping data. *Diabetes, Obesity & Metabolism,* 11 Suppl 1, pp.2-7.

Bruce, B. & Fries, J. F., 2003. The Stanford Health Assessment Questionnaire: dimensions and practical applications. *Health & Quality of Life Outcomes,* 1, pp.20.

Bruynesteyn, K., Van Der Heijde, D., Boers, M., Saudan, A., Peloso, P., Paulus, H., et al., 2002. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. *Arthritis Rheum,* 46, pp.913-20.

Buchs, N., Di Giovine, F. S., Silvestri, T., Vannier, E., Duff, G. W. & Miossec, P., 2001. IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: interaction with their plasma levels. *Genes Immun,* 2, pp.222-8.

Burmester, G. R., Blanco, R., Charles-Schoeman, C., Wollenhaupt, J., Zerbini, C., Benda, B., et al., 2013. Tofacitinib (CP-690,550) in combination with methotrexate

in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: a randomised phase 3 trial. *Lancet,* 381, pp.451-60.

Burmester, G. R., Stuhlmuller, B., Keyszer, G. & Kinne, R. W., 1997. Mononuclear phagocytes and rheumatoid synovitis. Mastermind or workhorse in arthritis? *Arthritis Rheum,* 40, pp.5-18.

Burns, P. 2004. *Perfomance Measurement via Random Portfolios* [Online]. Burns Statistics website. Available: http://www.burns-stat.com/pages/Working/perfmeasrandport.pdf [Accessed 15th April 2014].

Bykerk, V. P., 2011. Strategies to prevent rheumatoid arthritis in high-risk patients. *Curr Opin Rheumatol,* 23, pp.179-84.

Cantagrel, A., Navaux, F., Loubet-Lescoulie, P., Nourhashemi, F., Enault, G., Abbal, M., et al., 1999. Interleukin-1beta, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis. *Arthritis Rheum,* 42, pp.1093-100.

Capell, H. A., Porter, D. R., Madhok, R. & Hunter, J. A., 1993. Second line (disease modifying) treatment in rheumatoid arthritis: which drug for which patient? *Ann Rheum Dis,* 52, pp.423-8.

Carroll, M. C., 2004. The complement system in regulation of adaptive immunity. *Nat Immunol,* 5, pp.981-6.

Cerhan, J. R., Saag, K. G., Criswell, L. A., Merlino, L. A. & Mikuls, T. R., 2002. Blood transfusion, alcohol use, and anthropometric risk factors for rheumatoid arthritis in older women. *J Rheumatol,* 29, pp.246-54.

Chabchoub, G., Petit-Teixeira, E., Maalej, A., Pierlot, C., Bahloul, Z., Cornelis, F., et al., 2006. Lack of association between signaling lymphocytic activation molecule family member 1 gene and rheumatoid arthritis in the French and Tunisian populations. *Ann Rheum Dis,* 65, pp.1538-9.

Chabchoub, G., Teixiera, E. P., Maalej, A., Ben Hamad, M., Bahloul, Z., Cornelis, F., et al., 2009. The R620W polymorphism of the protein tyrosine phosphatase 22

gene in autoimmune thyroid diseases and rheumatoid arthritis in the Tunisian population. *Ann Hum Biol,* 36, pp.342-9.

Chibnik, L. B., Keenan, B. T., Cui, J., Liao, K. P., Costenbader, K. H., Plenge, R. M., et al., 2011. Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. *PLoS ONE,* 6, pp.e24380.

Choy, E. H. S., Smith, C. M., Farewell, V., Walker, D., Hassell, A., Chau, L., et al., 2008. Factorial randomised controlled trial of glucocorticoids and combination disease modifying drugs in early rheumatoid arthritis. *Ann Rheum Dis,* 67, pp.656-63.

Clavel, C., Nogueira, L., Laurent, L., Iobagiu, C., Vincent, C., Sebbag, M., et al., 2008. Induction of macrophage secretion of tumor necrosis factor alpha through Fcgamma receptor IIa engagement by rheumatoid arthritis-specific autoantibodies to citrullinated proteins complexed with fibrinogen. *Arthritis Rheum,* 58, pp.678-88.

Cohen, S. B., Emery, P., Greenwald, M. W., Dougados, M., Furie, R. A., Genovese, M. C., et al., 2006. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum,* 54, pp.2793-806.

Cojocaru, M., Cojocaru, I. M., Silosi, I., Vrabie, C. D. & Tanasescu, R., 2010. Extra-articular Manifestations in Rheumatoid Arthritis. *Medica,* 5, pp.286-91.

Cook, N. R., 2008. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem,* 54, pp.17-23.

Costenbader, K. H., Feskanich, D., Holmes, M., Karlson, E. W. & Benito-Garcia, E., 2008. Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women. *Ann Rheum Dis,* 67, pp.530-5.

Costenbader, K. H., Feskanich, D., Mandl, L. A. & Karlson, E. W., 2006. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med,* 119, pp.503.e1-9.

Crouch, D. J., Goddard, G. H. & Lewis, C. M., 2013. REGENT: a risk assessment and classification algorithm for genetic and environmental factors. *European Journal of Human Genetics,* 21, pp.109-11.

Crowson, C. S., Matteson, E. L., Davis, J. M., 3rd & Gabriel, S. E., 2013. Contribution of obesity to the rise in incidence of rheumatoid arthritis. *Arthritis Care Res,* 65, pp.71-7.

Cui, J., Stahl, E. A., Saevarsdottir, S., Miceli, C., Diogo, D., Trynka, G., et al., 2013. Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. *PLoS Genet,* 9, pp.e1003394.

Cui, J., Taylor, K. E., Lee, Y. C., Kallberg, H., Weinblatt, M. E., Coblyn, J. S., et al., 2014. The influence of polygenic risk scores on heritability of anti-CCP level in RA. *Genes Immun,* 15, pp.107-14.

D'agostino, R. B., Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., et al., 2008. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation,* 117, pp.743-53.

Daha, N. A. & Toes, R. E. M., 2011. Rheumatoid arthritis: Are ACPA-positive and ACPA-negative RA the same disease? *Nat Rev Rheumatol,* 7, pp.202-3.

Daly, A. K. & Day, C. P., 2001. Candidate gene case-control association studies: advantages and potential pitfalls. *Br J Clin Pharmacol,* 52, pp.489-99.

Davies, H. T., Crombie, I. K. & Tavakoli, M., 1998. When can odds ratios mislead? *BMJ,* 316, pp.989-91.

De Bakker, P. I., Ferreira, M. A., Jia, X., Neale, B. M., Raychaudhuri, S. & Voight, B. F., 2008. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet,* 17, pp.R122-8.

De Bakker, P. I., Mcvean, G., Sabeti, P. C., Miretti, M. M., Green, T., Marchini, J., et al., 2006. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet,* 38, pp.1166-72.

De Jager, P. L., Chibnik, L. B., Cui, J., Reischl, J., Lehr, S., Simon, K. C., et al., 2009. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol,* 8, pp.1111-9.

De Pablo, P., Dietrich, T. & Mcalindon, T. E., 2008. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. *J Rheumatol,* 35, pp.70-6.

De Rooy, D. P., Yeremenko, N. G., Wilson, A. G., Knevel, R., Lindqvist, E., Saxne, T., et al., 2013a. Genetic studies on components of the Wnt signalling pathway and the severity of joint destruction in rheumatoid arthritis. *Ann Rheum Dis,* 72, pp.769-75.

De Rooy, D. P., Zhernakova, A., Tsonaka, R., Willemze, A., Kurreeman, B. A., Trynka, G., et al., 2013b. A genetic variant in the region of MMP-9 is associated with serum levels and progression of joint damage in rheumatoid arthritis. *Ann Rheum Dis,* [Epub ahead of print].

De Vries-Bouwstra, J. K., Goekoop-Ruiterman, Y. P. M., Verpoort, K. N., Schreuder, G. M. T., Ewals, J. a. P. M., Terwiel, J. P., et al., 2008. Progression of joint damage in early rheumatoid arthritis: association with HLA-DRB1, rheumatoid factor, and anti-citrullinated protein antibodies in relation to different treatment strategies. *Arthritis Rheum,* 58, pp.1293-8.

Deal, C. L., Meenan, R. F., Goldenberg, D. L., Anderson, J. J., Sack, B., Pastan, R. S., et al., 1985. The clinical features of elderly-onset rheumatoid arthritis. A comparison with younger-onset disease of similar duration. *Arthritis Rheum,* 28, pp.987-94.

Deighton, C., O'mahony, R., Tosh, J., Turner, C., Rudolf, M. & Guideline Development Group, 2009. Management of rheumatoid arthritis: summary of NICE guidance. *BMJ,* 338, pp.b702.

Dervieux, T., Furst, D., Lein, D. O., Capps, R., Smith, K., Caldwell, J., et al., 2005. Pharmacogenetic and metabolite measurements are associated with clinical status in patients with rheumatoid arthritis treated with methotrexate: results of a multicentred cross sectional observational study. *Ann Rheum Dis,* 64, pp.1180-5.

Di Giuseppe, D., Alfredsson, L., Bottai, M., Askling, J. & Wolk, A., 2012. Long term alcohol intake and risk of rheumatoid arthritis in women: a population based cohort study. *BMJ,* 345, pp.e4230.

Dichamp, I., Bourgeois, A., Dirand, C., Herbein, G. & Wendling, D., 2007. Increased nuclear factor-kappaB activation in peripheral blood monocytes of patients with rheumatoid arthritis is mediated primarily by tumor necrosis factor-alpha. *J Rheumatol,* 34, pp.1976-83.

Dieye, A., Diallo, S., Diatta, M., Thiam, A., Ndiaye, R., Bao, O., et al., 1997. Identification of HLA-DR alleles for susceptibility to rheumatoid polyarthritis in Senegal. *Dakar Med,* 42, pp.111-3.

Ding, B., Padyukov, L., Lundstrom, E., Seielstad, M., Plenge, R. M., Oksenberg, J. R., et al., 2009. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis Rheum,* 60, pp.30-8.

Diogo, D., Kurreeman, F., Stahl, E. A., Liao, K. P., Gupta, N., Greenberg, J. D., et al., 2013. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *Am J Hum Genet,* 92, pp.15-27.

Dissick, A., Redman, R. S., Jones, M., Rangan, B. V., Reimold, A., Griffiths, G. R., et al., 2010. Association of periodontitis with rheumatoid arthritis: a pilot study. *J Periodontol,* 81, pp.223-30.

Drossaers-Bakker, K. W., De Buck, M., Van Zeben, D., Zwinderman, A. H., Breedveld, F. C. & Hazes, J. M., 1999. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. *Arthritis Rheum,* 42, pp.1854-60.

Drossaers-Bakker, K. W., Zwinderman, A. H., Vliet Vlieland, T. P. M., Van Zeben, D., Vos, K., Breedveld, F. C., et al., 2002. Long-term outcome in rheumatoid arthritis: a simple algorithm of baseline parameters can predict radiographic damage, disability, and disease course at 12-year followup. *Arthritis Rheum,* 47, pp.383-90.

Drouin, J., Haraoui, B. & E Initiative Group, 2010. Predictors of clinical response and radiographic progression in patients with rheumatoid arthritis treated with methotrexate monotherapy. *J Rheumatol,* 37, pp.1405-10.

Du Montcel, S. T., Michou, L., Petit-Teixeira, E., Osorio, J., Lemaire, I., Lasbleiz, S., et al., 2005. New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum,* 52, pp.1063-8.

Dupont, C., Armant, D. R. & Brenner, C. A., 2009. Epigenetics: definition, mechanisms and clinical perspective. *Semin Reprod Med,* 27, pp.351-7.

Ellinghaus, D., Ellinghaus, E., Nair, R. P., Stuart, P. E., Esko, T., Metspalu, A., et al., 2012. Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. *Am J Hum Genet,* 90, pp.636-47.

Emery, P., Durez, P., Dougados, M., Legerton, C. W., Becker, J. C., Vratsanos, G., et al., 2010. Impact of T-cell costimulation modulation in patients with undifferentiated inflammatory arthritis or very early rheumatoid arthritis: a clinical and imaging study of abatacept (the ADJUST trial). *Ann Rheum Dis,* 69, pp.510-6.

Entezami, P., Fox, D. A., Clapham, P. J. & Chung, K. C., 2011. Historical perspective on the etiology of rheumatoid arthritis. *Hand Clin,* 27, pp.1-10.

Eras Study Group, 2000. Socioeconomic deprivation and rheumatoid disease: what lessons for the health service? ERAS Study Group. Early Rheumatoid Arthritis Study. *Ann Rheum Dis,* 59, pp.794-9.

Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., et al., 2012. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet,* 44, pp.1336-40.

Feitsma, A. L., Van Der Helm-Van Mil, A. H. M., Huizinga, T. W. J., De Vries, R. R. P. & Toes, R. E. M., 2008. Protection against rheumatoid arthritis by HLA: nature and nurture. *Ann Rheum Dis,* 67 Suppl 3, pp.iii61-3.

Feitsma, A. L., Van Der Voort, E. I. H., Franken, K. L. M. C., El Bannoudi, H., Elferink, B. G., Drijfhout, J. W., et al., 2010. Identification of citrullinated vimentin

peptides as T cell epitopes in HLA-DR4-positive patients with rheumatoid arthritis. *Arthritis Rheum,* 62, pp.117-25.

Firestein, G. S., Alvaro-Gracia, J. M. & Maki, R., 1990. Quantitative analysis of cytokine gene expression in rheumatoid arthritis. *J Immunol,* 144, pp.3347-53.

Fleischmann, R., Kremer, J., Cush, J., Schulze-Koops, H., Connell, C. A., Bradley, J. D., et al., 2012. Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N Engl J Med,* 367, pp.495-507.

Fransen, J., Kooloos, W. M., Wessels, J. A., Huizinga, T. W., Guchelaar, H. J., Van Riel, P. L., et al., 2012. Clinical pharmacogenetic model to predict response of MTX monotherapy in patients with established rheumatoid arthritis after DMARD failure. *Pharmacogenomics,* 13, pp.1087-94.

Fries, J. F., Hochberg, M. C., Medsger, T. A., Jr., Hunder, G. G. & Bombardier, C., 1994. Criteria for rheumatic disease. Different types and different functions. The American College of Rheumatology Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum,* 37, pp.454-62.

Frisell, T., Holmqvist, M., Kallberg, H., Klareskog, L., Alfredsson, L. & Askling, J., 2013. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis Rheum,* 65, pp.2773-82.

Garnero, P., Gineyts, E., Christgau, S., Finck, B. & Delmas, P. D., 2002a. Association of baseline levels of urinary glucosyl-galactosyl-pyridinoline and type II collagen C-telopeptide with progression of joint destruction in patients with early rheumatoid arthritis. *Arthritis Rheum,* 46, pp.21-30.

Garnero, P., Landewe, R., Boers, M., Verhoeven, A., Van Der Linden, S., Christgau, S., et al., 2002b. Association of baseline levels of markers of bone and cartilage degradation with long-term progression of joint damage in patients with early rheumatoid arthritis: the COBRA study. *Arthritis Rheum,* 46, pp.2847-56.

Garrood, T., Shattles, W. & Scott, D. L., 2011. Treating early rheumatoid arthritis intensively: current UK practice does not reflect guidelines. *Clin Rheumatol,* 30, pp.103-6.

Genovese, M. C., Becker, J.-C., Schiff, M., Luggen, M., Sherrer, Y., Kremer, J., et al., 2005. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. *N Engl J Med,* 353, pp.1114-23.

Gerlag, D. M., Raza, K., Van Baarsen, L. G. M., Brouwer, E., Buckley, C. D., Burmester, G. R., et al., 2012. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. *Ann Rheum Dis,* 71, pp.638-41.

Gibson, G., 2011. Rare and common variants: twenty arguments. *Nat Rev Genet,* 13, pp.135-45.

Ginanjar, E., Sumariyono, Setiati, S. & Setiyohadi, B., 2007. Vitamin D and autoimmune disease. *Acta Med Indones,* 39, pp.133-41.

Goekoop-Ruiterman, Y. P. M., De Vries-Bouwstra, J. K., Allaart, C. F., Van Zeben, D., Kerstens, P. J. S. M., Hazes, J. M. W., et al., 2008. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): A randomized, controlled trial. *Arthritis Rheum,* 58, pp.S126-35.

Gomez, R., Conde, J., Scotece, M., Gomez-Reino, J. J., Lago, F. & Gualillo, O., 2011. What's new in our understanding of the role of adipokines in rheumatic diseases? *Nat Rev Rheumatol,* 7, pp.528-36.

Gonzalez-Gay, M. A., Garcia-Porrua, C. & Hajeer, A. H., 2002. Influence of human leukocyte antigen-DRB1 on the susceptibility and severity of rheumatoid arthritis. *Semin Arthritis Rheum,* 31, pp.355-60.

Gorman, J. D., Lum, R. F., Chen, J. J., Suarez-Almazor, M. E., Thomson, G. & Criswell, L. A., 2004. Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. *Arthritis Rheum,* 50, pp.400-12.

Gossec, L., Dougados, M., Goupille, P., Cantagrel, A., Sibilia, J., Meyer, O., et al., 2004. Prognostic factors for remission in early rheumatoid arthritis: a multiparameter prospective study. *Ann Rheum Dis,* 63, pp.675-80.

Gottenberg, J. E., Ravaud, P., Cantagrel, A., Combe, B., Flipo, R. M., Schaeverbeke, T., et al., 2012. Positivity for anti-cyclic citrullinated peptide is associated with a better response to abatacept: data from the 'Orencia and Rheumatoid Arthritis' registry. *Ann Rheum Dis,* 71, pp.1815-9.

Gourraud, P.-A., Boyer, J.-F., Barnetche, T., Abbal, M., Cambon-Thomsen, A., Cantagrel, A., et al., 2006. A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity. *Arthritis Rheum,* 54, pp.593-9.

Graudal, N. & Jurgens, G., 2010. Similar effects of disease-modifying antirheumatic drugs, glucocorticoids, and biologic agents on radiographic progression in rheumatoid arthritis: meta-analysis of 70 randomized placebo-controlled or drug-controlled studies, including 112 comparisons. *Arthritis Rheum,* 62, pp.2852-63.

Gregersen, P. K., Silver, J. & Winchester, R. J., 1987. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum,* 30, pp.1205-13.

Guo, Y., Logan, H. L., Glueck, D. H. & Muller, K. E., 2013. Selecting a sample size for studies with repeated measures. *BMC Med Res Methodol,* 13, pp.100.

Guthrie, K. A., Dugowson, C. E., Voigt, L. F., Koepsell, T. D. & Nelson, J. L., 2010. Does pregnancy provide vaccine-like protection against rheumatoid arthritis? *Arthritis Rheum,* 62, pp.1842-8.

Han, B., Diogo, D., Eyre, S., Kallberg, H., Zhernakova, A., Bowes, J., et al., 2014. Fine Mapping Seronegative and Seropositive Rheumatoid Arthritis to Shared and Distinct HLA Alleles by Adjusting for the Effects of Heterogeneity. *Am J Hum Genet,* [Epub ahead of print].

Harrison, B. J., Silman, A. J. & Symmons, D. P., 2000. Does the age of onset of rheumatoid arthritis influence phenotype?: a prospective study of outcome and prognostic factors. *Rheumatology (Oxford),* 39, pp.112-3.

Hashimoto, J., Garnero, P., Van Der Heijde, D., Miyasaka, N., Yamamoto, K., Kawai, S., et al., 2009. A combination of biochemical markers of cartilage and bone turnover, radiographic damage and body mass index to predict the progression of

joint destruction in patients with rheumatoid arthritis treated with disease-modifying anti-rheumatic drugs. *Mod Rheumatol,* 19, pp.273-82.

Hazes, J. M., Dijkmans, B. A., Vandenbroucke, J. P., De Vries, R. R. & Cats, A., 1990. Pregnancy and the risk of developing rheumatoid arthritis. *Arthritis Rheum,* 33, pp.1770-5.

Heliovaara, M., Aho, K., Knekt, P., Impivaara, O., Reunanen, A. & Aromaa, A., 2000. Coffee consumption, rheumatoid factor, and the risk of rheumatoid arthritis. *Ann Rheum Dis,* 59, pp.631-5.

Heliovaara, M., Aho, K., Reunanen, A., Knekt, P. & Aromaa, A., 1995. Parity and risk of rheumatoid arthritis in Finnish women. *Br J Rheumatol,* 34, pp.625-8.

Helliwell, P. S. & Ibrahim, G., 2003. Ethnic differences in responses to disease modifying drugs. *Rheumatology (Oxford),* 42, pp.1197-201.

Hernandez Avila, M., Liang, M. H., Willett, W. C., Stampfer, M. J., Colditz, G. A., Rosner, B., et al., 1990. Reproductive factors, smoking, and the risk for rheumatoid arthritis. *Epidemiology,* 1, pp.285-91.

Hetland, M. L., Ejbjerg, B., Horslev-Petersen, K., Jacobsen, S., Vestergaard, A., Jurik, A. G., et al., 2009. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann Rheum Dis,* 68, pp.384-90.

Hider, S. L., Buckley, C., Silman, A. J., Symmons, D. P. M. & Bruce, I. N., 2005. Factors influencing response to disease modifying antirheumatic drugs in patients with rheumatoid arthritis. *J Rheumatol,* 32, pp.11-16.

Higgins, J. P. T. & Green, S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions* [Online]. Available: http://handbook.cochrane.org/chapter_16/16_1_2_general_principles_for_dealing_with_missing_data.htm [Accessed 21st April 2014].

Hill, J. A., Bell, D. A., Brintnell, W., Yue, D., Wehrli, B., Jevnikar, A. M., et al., 2008. Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in DR4-IE transgenic mice. *J Exp Med,* 205, pp.967-79.

Hill, J. A., Southwood, S., Sette, A., Jevnikar, A. M., Bell, D. A. & Cairns, E., 2003. Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J Immunol,* 171, pp.538-41.

Hodkinson, B., Musenge, E., Ally, M., Meyer, P. W., Anderson, R. & Tikly, M., 2012. Response to traditional disease-modifying anti-rheumatic drugs in indigent South Africans with early rheumatoid arthritis. *Clin Rheumatol,* 31, pp.613-9.

Howie, B. N., Donnelly, P. & Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet,* 5, pp.e1000529.

Hughes, L. B., Morrison, D., Kelley, J. M., Padilla, M. A., Vaughan, L. K., Westfall, A. O., et al., 2008. The HLA-DRB1 shared epitope is associated with susceptibility to rheumatoid arthritis in African Americans through European genetic admixture. *Arthritis Rheum,* 58, pp.349-58.

Huizinga, T. W., Keijsers, V., Yanni, G., Hall, M., Ramage, W., Lanchbury, J., et al., 2000. Are differences in interleukin 10 production associated with joint damage? *Rheumatology (Oxford),* 39, pp.1180-8.

Hulsmans, H. M., Jacobs, J. W., Van Der Heijde, D. M., Van Albada-Kuipers, G. A., Schenk, Y. & Bijlsma, J. W., 2000. The course of radiologic damage during the first six years of rheumatoid arthritis. *Arthritis Rheum,* 43, pp.1927-40.

Hyrich, K. L., Watson, K. D., Silman, A. J., Symmons, D. P. & British Society for Rheumatology Biologics Register, 2006. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. *Rheumatology (Oxford),* 45, pp.1558-65.

Intlekofer, A. M., Takemoto, N., Wherry, E. J., Longworth, S. A., Northrup, J. T., Palanivel, V. R., et al., 2005. Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat Immunol,* 6, pp.1236-44.

Irigoyen, P., Lee, A. T., Wener, M. H., Li, W., Kern, M., Batliwalla, F., et al., 2005. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis:

contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis Rheum,* 52, pp.3813-8.

Isaacs, J. D., Cohen, S. B., Emery, P., Tak, P. P., Wang, J., Lei, G., et al., 2013. Effect of baseline rheumatoid factor and anticitrullinated peptide antibody serotype on rituximab clinical response: a meta-analysis. *Ann Rheum Dis,* 72, pp.329-36.

Jaenisch, R. & Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet,* 33 Suppl, pp.245-54.

Jain, M. & Samuels, J., 2011. Musculoskeletal ultrasound as a diagnostic and prognostic tool in rheumatoid arthritis. *Bull NYU Hosp Jt Dis,* 69, pp.215-9.

Janssens, A. C. J. W., Moonesinghe, R., Yang, Q., Steyerberg, E. W., Van Duijn, C. M. & Khoury, M. J., 2007. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med,* 9, pp.528-35.

Jawaheer, D., Maranian, P., Park, G., Lahiff, M., Amjadi, S. S. & Paulus, H. E., 2010. Disease progression and treatment responses in a prospective DMARD-naive seropositive early rheumatoid arthritis cohort: does gender matter? *J Rheumatol,* 37, pp.2475-85.

Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.-M., Concannon, P. J., Rich, S. S., et al., 2013. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE,* 8, pp.e64683.

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'donnell, C. J. & De Bakker, P. I. W., 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics,* 24, pp.2938-9.

Jonsson, I.-M., Verdrengh, M., Brisslert, M., Lindblad, S., Bokarewa, M., Islander, U., et al., 2007. Ethanol prevents development of destructive arthritis. *Proc Natl Acad Sci U S A,* 104, pp.258-63.

Jonsson, T., Arinbjarnarson, S., Thorsteinsson, J., Steinsson, K., Geirsson, A. J., Jonsson, H., et al., 1995. Raised IgA rheumatoid factor (RF) but not IgM RF or IgG

RF is associated with extra-articular manifestations in rheumatoid arthritis. *Scand J Rheumatol,* 24, pp.372-5.

Jonsson, T. & Valdimarsson, H., 1998. What about IgA rheumatoid factor in rheumatoid arthritis? *Ann Rheum Dis,* 57, pp.63-4.

Jorgensen, C., Picot, M. C., Bologna, C. & Sany, J., 1996. Oral contraception, parity, breast feeding, and severity of rheumatoid arthritis. *Ann Rheum Dis,* 55, pp.94-8.

Jorgensen, K. T., Pedersen, B. V., Jacobsen, S., Biggar, R. J. & Frisch, M., 2010. National cohort study of reproductive risk factors for rheumatoid arthritis in Denmark: a role for hyperemesis, gestational hypertension and pre-eclampsia? *Ann Rheum Dis,* 69, pp.358-63.

Kalla, A. A. & Tikly, M., 2003. Rheumatoid arthritis in the developing world. *Best Pract Res Clin Rheumatol,* 17, pp.863-75.

Kallberg, H., Jacobsen, S., Bengtsson, C., Pedersen, M., Padyukov, L., Garred, P., et al., 2009. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. *Ann Rheum Dis,* 68, pp.222-7.

Kallberg, H., Padyukov, L., Plenge, R. M., Ronnelid, J., Gregersen, P. K., Van Der Helm-Van Mil, A. H. M., et al., 2007. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. *Am J Hum Genet,* 80, pp.867-75.

Kaneko, Y., Kuwana, M., Kameda, H. & Takeuchi, T., 2011. Sensitivity and specificity of 2010 rheumatoid arthritis classification criteria. *Rheumatology (Oxford),* 50, pp.1268-74.

Karlson, E. W., Chibnik, L. B., Cui, J., Plenge, R. M., Glass, R. J., Maher, N. E., et al., 2008. Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study. *Ann Rheum Dis,* 67, pp.358-63.

Karlson, E. W., Chibnik, L. B., Kraft, P., Cui, J., Keenan, B. T., Ding, B., et al., 2010. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. *Ann Rheum Dis,* 69, pp.1077-1085.

Karlson, E. W., Ding, B., Keenan, B. T., Liao, K., Costenbader, K. H., Klareskog, L., et al., 2013. Association of environmental and genetic factors and gene-environment interactions with risk of developing rheumatoid arthritis. *Arthritis Care Res,* 65, pp.1147-56.

Karlson, E. W., Mandl, L. A., Aweh, G. N. & Grodstein, F., 2003. Coffee consumption and risk of rheumatoid arthritis. *Arthritis Rheum,* 48, pp.3055-60.

Karlson, E. W., Mandl, L. A., Hankinson, S. E. & Grodstein, F., 2004. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. *Arthritis Rheum,* 50, pp.3458-67.

Kawabe, T., Naka, T., Yoshida, K., Tanaka, T., Fujiwara, H., Suematsu, S., et al., 1994. The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity,* 1, pp.167-78.

Kayakabe, K., Kuroiwa, T., Sakurai, N., Ikeuchi, H., Kadiombo, A. T., Sakairi, T., et al., 2012. Interleukin-1 measurement in stimulated whole blood cultures is useful to predict response to anti-TNF therapies in rheumatoid arthritis. *Rheumatology (Oxford),* 51, pp.1639-43.

Klareskog, L., Padyukov, L., Ronnelid, J. & Alfredsson, L., 2006a. Genes, environment and immunity in the development of rheumatoid arthritis. *Curr Opin Immunol,* 18, pp.650-5.

Klareskog, L., Stolt, P., Lundberg, K., Kallberg, H., Bengtsson, C., Grunewald, J., et al., 2006b. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum,* 54, pp.38-46.

Knevel, R., De Rooy, D. P., Gregersen, P. K., Lindqvist, E., Wilson, A. G., Grondal, G., et al., 2012a. Studying associations between variants in TRAF1-C5 and TNFAIP3-OLIG3 and the progression of joint destruction in rheumatoid arthritis in multiple cohorts. *Ann Rheum Dis,* 71, pp.1753-5.

Knevel, R., De Rooy, D. P. C., Zhernakova, A., Grondal, G., Krabben, A., Steinsson, K., et al., 2013a. Association of variants in IL2RA with progression of joint destruction in rheumatoid arthritis. Arthritis Rheum. 65, pp.1684-93.

Knevel, R., Grondal, G., Huizinga, T. W. J., Visser, A. W., Jonsson, H., Vikingsson, A., et al., 2012b. Genetic predisposition of the severity of joint destruction in rheumatoid arthritis: a population-based study. *Ann Rheum Dis,* 71, pp.707-9.

Knevel, R., Klein, K., Somers, K., Ospelt, C., Houwing-Duistermaat, J., Van Nies, J., et al., 2013b. Identification of a genetic variant for joint damage progression in autoantibody-positive rheumatoid arthritis. *Ann Rheum Dis,* [Epub ahead of print].

Knevel, R., Krabben, A., Brouwer, E., Posthumus, M. D., Wilson, A. G., Lindqvist, E., et al., 2012a. Genetic variants in IL15 associate with progression of joint destruction in rheumatoid arthritis: a multicohort study. *Ann Rheum Dis,* 71, pp.1651-7.

Knevel, R., Krabben, A., Wilson, A. G., Brouwer, E., Leijsma, M. K., Lindqvist, E., et al., 2013a. A genetic variant in granzyme B is associated with progression of joint destruction in rheumatoid arthritis. *Arthritis Rheum,* 65, pp.582-9.

Kotake, S., Udagawa, N., Takahashi, N., Matsuzaki, K., Itoh, K., Ishiyama, S., et al., 1999. IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. *J Clin Invest,* 103, pp.1345-52.

Krabben, A., Wilson, A. G., De Rooy, D. P. C., Zhernakova, A., Brouwer, E., Lindqvist, E., et al., 2013. Association of genetic variants in the IL4 and IL4R genes with the severity of joint damage in rheumatoid arthritis: a study in seven cohorts. *Arthritis Rheum,* 65, pp.3051-7.

Kristensen, L. E., Kapetanovic, M. C., Gulfe, A., Soderlin, M., Saxne, T. & Geborek, P., 2008. Predictors of response to anti-TNF therapy according to ACR and EULAR criteria in patients with established RA: results from the South Swedish Arthritis Treatment Group Register. *Rheumatology (Oxford),* 47, pp.495-9.

Kuiper, S., Van Gestel, A. M., Swinkels, H. L., De Boo, T. M., Da Silva, J. A. & Van Riel, P. L., 2001. Influence of sex, age, and menopausal state on the course of early rheumatoid arthritis. *J Rheumatol,* 28, pp.1809-16.

Kurreeman, F., Liao, K., Chibnik, L., Hickey, B., Stahl, E., Gainer, V., et al., 2011. Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am J Hum Genet,* 88, pp.57-69.

Kurreeman, F. a. S., Padyukov, L., Marques, R. B., Schrodi, S. J., Seddighzadeh, M., Stoeken-Rijsbergen, G., et al., 2007. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med,* 4, pp.e278.

Lahiri, M., Luben, R. N., Morgan, C., Bunn, D. K., Marshall, T., Lunt, M., et al., 2014. Using lifestyle factors to identify individuals at higher risk of inflammatory polyarthritis (results from the European Prospective Investigation of Cancer-Norfolk and the Norfolk Arthritis Register--the EPIC-2-NOAR Study). *Ann Rheum Dis,* 73, pp.219-26.

Lajas, C., Abasolo, L., Bellajdel, B., Hernandez-Garcia, C., Carmona, L., Vargas, E., et al., 2003. Costs and predictors of costs in rheumatoid arthritis: a prevalence-based study. *Arthritis & Rheumatism,* 49, pp.64-70.

Landewe, R. B. M., Boers, M., Verhoeven, A. C., Westhovens, R., Van De Laar, M. a. F. J., Markusse, H. M., et al., 2002. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. *Arthritis Rheum,* 46, pp.347-56.

Lankarani-Fard, A., Kritz-Silverstein, D., Barrett-Connor, E. & Goodman-Gruen, D., 2001. Cumulative duration of breast-feeding influences cortisol levels in postmenopausal women. *J Womens Health Gend Based Med,* 10, pp.681-7.

Lawrence, J. S., 1961. Prevalence of rheumatoid arthritis. *Ann Rheum Dis,* 20, pp.11-7.

Lemeshow, S. & Hosmer, D. W., Jr., 1982. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol,* 115, pp.92-106.

Lindqvist, E., Eberhardt, K., Bendtzen, K., Heinegard, D. & Saxne, T., 2005. Prognostic laboratory markers of joint damage in rheumatoid arthritis. *Ann Rheum Dis,* 64, pp.196-201.

Lindqvist, E., Jonsson, K., Saxne, T. & Eberhardt, K., 2003. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. *Ann Rheum Dis,* 62, pp.611-6.

Liu, C., Batliwalla, F., Li, W., Lee, A., Roubenoff, R., Beckman, E., et al., 2008. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol Med,* 14, pp.575-81.

Liu, Y., Aryee, M. J., Padyukov, L., Fallin, M. D., Hesselberg, E., Runarsson, A., et al., 2013. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol,* 31, pp.142-7.

Lu, B., Solomon, D. H., Costenbader, K. H. & Karlson, E. W., 2014. Alcohol consumption and risk of incident rheumatoid arthritis in women: A prospective study. *Arthritis Rheum,* [Epub ahead of print].

Lu, Q. & Elston, R. C., 2008. Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes. *Am J Hum Genet,* 82, pp.641-651.

Lunetta, K. L., 2008. Genetic association studies. *Circulation,* 118, pp.96-101.

Ma, M. H. Y., Scott, I. C., Dahanayake, C., Cope, A. P. & Scott, D. L., 2014. Clinical And Serological Predictors Of Remission In Rheumatoid Arthritis Are Dependent On Treatment Regimes. *J Rheumatol,* In-Press.

Ma, M. H. Y., Scott, I. C., Kingsley, G. H. & Scott, D. L., 2010. Remission in early rheumatoid arthritis. *J Rheumatol,* 37, pp.1444-53.

Makrygiannakis, D., Hermansson, M., Ulfgren, A. K., Nicholas, A. P., Zendman, A. J. W., Eklund, A., et al., 2008. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis,* 67, pp.1488-92.

Malemba, J. J., Mbuyi-Muamba, J. M., Mukaya, J., Bossuyt, X., Verschueren, P. & Westhovens, R., 2012. The epidemiology of rheumatoid arthritis in Kinshasa,

Democratic Republic of Congo--a population-based study. *Rheumatology (Oxford)*, 51, pp.1644-7.

Malik, F. & Ranganathan, P., 2013. Methotrexate pharmacogenetics in rheumatoid arthritis: a status report. *Pharmacogenomics*, 14, pp.305-14.

Mancarella, L., Bobbio-Pallavicini, F., Ceccarelli, F., Falappone, P. C., Ferrante, A., Malesci, D., et al., 2007. Good clinical response, remission, and predictors of remission in rheumatoid arthritis patients treated with tumor necrosis factor-alpha blockers: the GISEA study. *J Rheumatol*, 34, pp.1670-3.

Mandrekar, P., Catalano, D., White, B. & Szabo, G., 2006. Moderate alcohol intake in humans attenuates monocyte inflammatory responses: inhibition of nuclear regulatory factor kappa B and induction of interleukin 10. *Alcohol Clin Exp Res*, 30, pp.135-9.

Manfredsdottir, V. F., Vikingsdottir, T., Jonsson, T., Geirsson, A. J., Kjartansson, O., Heimisdottir, M., et al., 2006. The effects of tobacco smoking and rheumatoid factor seropositivity on disease activity and joint damage in early rheumatoid arthritis. *Rheumatology (Oxford)*, 45, pp.734-40.

Mangat, P., Wegner, N., Venables, P. J. & Potempa, J., 2010. Bacterial and human peptidylarginine deiminases: targets for inhibiting the autoimmune response in rheumatoid arthritis? *Arthritis Res Ther*, 12, pp.209.

Marinou, I., Healy, J., Mewar, D., Moore, D. J., Dickson, M. C., Binks, M. H., et al., 2007. Association of interleukin-6 and interleukin-10 genotypes with radiographic damage in rheumatoid arthritis is dependent on autoantibody status. *Arthritis Rheum*, 56, pp.2549-56.

Marinou, I., Maxwell, J. R. & Wilson, A. G., 2010. Genetic influences modulating the radiological severity of rheumatoid arthritis. *Ann Rheum Dis*, 69, pp.476-82.

Marteau, T. M., French, D. P., Griffin, S. J., Prevost, A. T., Sutton, S., Watkinson, C., et al., 2010. Effects of communicating DNA-based disease risk estimates on risk-reducing behaviours. *Cochrane Database Syst Rev*, pp.CD007275.

Martinez-Gamboa, L., Brezinschek, H.-P., Burmester, G. R. & Dorner, T., 2006. Immunopathologic role of B lymphocytes in rheumatoid arthritis: rationale of B cell-directed therapy. *Autoimmun Rev,* 5, pp.437-42.

Masdottir, B., Jonsson, T., Manfredsdottir, V., Vikingsson, A., Brekkan, A. & Valdimarsson, H., 2000. Smoking, rheumatoid factor isotypes and severity of rheumatoid arthritis. *Rheumatology (Oxford),* 39, pp.1202-5.

Maxwell, J. R., Gowers, I. R., Moore, D. J. & Wilson, A. G., 2010. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. *Rheumatology (Oxford),* 49, pp.2140-6.

Mcentegart, A., Morrison, E., Capell, H. A., Duncan, M. R., Porter, D., Madhok, R., et al., 1997. Effect of social deprivation on disease severity and outcome in patients with rheumatoid arthritis. *Ann Rheum Dis,* 56, pp.410-3.

Mcqueen, F. M., Benton, N., Perry, D., Crabbe, J., Robinson, E., Yeoman, S., et al., 2003. Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. *Arthritis Rheum,* 48, pp.1814-27.

Medical Research Council. 2014. MRC Website. Available: http://www.mrc.ac.uk/Ourresearch/ResearchInitiatives/StratifiedMedicine/index.htm [Accessed 19th April 2014].

Mercado, F. B., Marshall, R. I., Klestov, A. C. & Bartold, P. M., 2001. Relationship between rheumatoid arthritis and periodontitis. *J Periodontol,* 72, pp.779-87.

Merlino, L. A., Cerhan, J. R., Criswell, L. A., Mikuls, T. R. & Saag, K. G., 2003. Estrogen and other female reproductive risk factors are not strongly associated with the development of rheumatoid arthritis in elderly women. *Semin Arthritis Rheum,* 33, pp.72-82.

Merlino, L. A., Curtis, J., Mikuls, T. R., Cerhan, J. R., Criswell, L. A., Saag, K. G., et al., 2004. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum,* 50, pp.72-7.

Metz, C. E., 1978. Basic principles of ROC analysis. *Semin Nucl Med,* 8, pp.283-298.

Mewar, D., Marinou, I., Coote, A. L., Moore, D. J., Akil, M., Smillie, D., et al., 2008. Association between radiographic severity of rheumatoid arthritis and shared epitope alleles: differing mechanisms of susceptibility and protection. *Ann Rheum Dis,* 67, pp.980-3.

Mikuls, T. R., Cerhan, J. R., Criswell, L. A., Merlino, L., Mudano, A. S., Burma, M., et al., 2002. Coffee, tea, and caffeine consumption and risk of rheumatoid arthritis: results from the Iowa Women's Health Study. *Arthritis Rheum,* 46, pp.83-91.

Mikuls, T. R., Sayles, H., Yu, F., Levan, T., Gould, K. A., Thiele, G. M., et al., 2010. Associations of cigarette smoking with rheumatoid arthritis in African Americans. *Arthritis Rheum,* 62, pp.3560-8.

Mjaavatten, M. D. & Bykerk, V. P., 2013. Early rheumatoid arthritis: The performance of the 2010 ACR/EULAR criteria for diagnosing RA. *Best Pract Res Clin Rheumatol,* 27, pp.451-66.

Mohamed, R. H., Pasha, H. F. & El-Shahawy, E. E., 2012. Influence of TRAF1/C5 and STAT4 genes polymorphisms on susceptibility and severity of rheumatoid arthritis in Egyptian population. *Cell Immunol,* 273, pp.67-72.

Mohammed, F. F., Smookler, D. S. & Khokha, R., 2003. Metalloproteinases, inflammation, and rheumatoid arthritis. *Ann Rheum Dis,* 62 Suppl 2, pp.ii43-7.

Moodley, D., Mody, G. M. & Chuturgoon, A. A., 2010. Functional analysis of the p53 codon 72 polymorphism in black South Africans with rheumatoid arthritis--a pilot study. *Clin Rheumatol,* 29, pp.1099-105.

Moons, K. G. M., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., et al., 2012. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart,* 98, pp.683-90.

Mottonen, T., Hannonen, P., Leirisalo-Repo, M., Nissila, M., Kautiainen, H., Korpela, M., et al., 1999. Comparison of combination therapy with single-drug

therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. *Lancet*, 353, pp.1568-73.

Myasoedova, E., Crowson, C. S., Kremers, H. M., Therneau, T. M. & Gabriel, S. E., 2010. Is the incidence of rheumatoid arthritis rising?: results from Olmsted County, Minnesota, 1955-2007. *Arthritis Rheum,* 62, pp.1576-82.

Myasoedova, E., Crowson, C. S., Turesson, C., Gabriel, S. E. & Matteson, E. L., 2011. Incidence of extraarticular rheumatoid arthritis in Olmsted County, Minnesota, in 1995-2007 versus 1985-1994: a population-based study. *J Rheumatol,* 38, pp.983-9.

Nakano, K., Whitaker, J. W., Boyle, D. L., Wang, W. & Firestein, G. S., 2013. DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis,* 72, pp.110-7.

Naredo, E., Collado, P., Cruz, A., Palop, M. J., Cabero, F., Richi, P., et al., 2007. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. *Arthritis Rheum,* 57, pp.116-24.

National Audit Office. 2009. *Services for people with rheumatoid arthritis* [Online]. Available: http://www.nao.org.uk/wp-content/uploads/2009/07/0809823es.pdf [Accessed 7th April 2014].

National Institute for Health and Care Excellence. 2007. *Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis* [Online]. NICE Website. Available: http://www.nice.org.uk/TA130 [Accessed 20th April 2014].

National Institute for Health and Care Excellence. 2010. *TA195 Rheumatoid arthritis - drugs for treatment after failure of a TNF inhibitor* [Online]. NICE Website. Available: http://guidance.nice.org.uk/TA195/QuickRefGuide/pdf/English [Accessed 20th April 2014].

Nelson, S. C., Doheny, K. F., Pugh, E. W., Romm, J. M., Ling, H., Laurie, C. A., et al., 2013. Imputation-based genomic coverage assessments of current human genotyping arrays. *G3 (Bethesda),* 3, pp.1795-807.

Nielen, M. M. J., Van Schaardenburg, D., Reesink, H. W., Van De Stadt, R. J., Van Der Horst-Bruinsma, I. E., De Koning, M. H. M. T., et al., 2004. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum,* 50, pp.380-6.

Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G., Nakazawa, T., Kawano, S., et al., 2007. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Ann Intern Med,* 146, pp.797-808.

Nissen, M. J., Gabay, C., Scherer, A., Finckh, A. & Swiss Clinical Quality Management Project in Rheumatoid, A., 2010. The effect of alcohol on radiographic progression in rheumatoid arthritis. *Arthritis Rheum,* 62, pp.1265-72.

Okada, Y., Wu, D., Trynka, G., Towfique, R., Chikashi, T., Katsunori, I., et al., 2013. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nat Genet,* 506, pp.376-81.

Ortea, I., Roschitzki, B., Ovalles, J. G., Longo, J. L., De La Torre, I., Gonzalez, I., et al., 2012. Discovery of serum proteomic biomarkers for prediction of response to infliximab (a monoclonal anti-TNF antibody) treatment in rheumatoid arthritis: an exploratory analysis. *J Proteomics,* 77, pp.372-82.

Padyukov, L., Seielstad, M., Ong, R. T. H., Ding, B., Ronnelid, J., Seddighzadeh, M., et al., 2011. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann Rheum Dis,* 70, pp.259-65.

Padyukov, L., Silva, C., Stolt, P., Alfredsson, L. & Klareskog, L., 2004. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum,* 50, pp.3085-92.

Paglieroni, T. G., Ward, J. & Holland, P. V., 1995. Changes in peripheral blood CD5 (Bla) B-cell populations and autoantibodies following blood transfusion. *Transfusion,* 35, pp.189-98.

Palosaari, K., Vuotila, J., Takalo, R., Jartti, A., Niemela, R. K., Karjalainen, A., et al., 2006. Bone oedema predicts erosive progression on wrist MRI in early RA--a 2-yr observational MRI and NC scintigraphy study. *Rheumatology (Oxford),* 45, pp.1542-8.

Parkes, M., Cortes, A., Van Heel, D. A. & Brown, M. A., 2013. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet,* 14, pp.661-73.

Pawlik, A., Kurzawski, M., Florczak, M., Gawronska Szklarz, B. & Herczynska, M., 2005. IL1beta+3953 exon 5 and IL-2 -330 promoter polymorphisms in patients with rheumatoid arthritis. *Clin Exp Rheumatol,* 23, pp.159-64.

Pearce, E. L., Mullen, A. C., Martins, G. A., Krawczyk, C. M., Hutchins, A. S., Zediak, V. P., et al., 2003. Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science,* 302, pp.1041-3.

Pedersen, M., Jacobsen, S., Garred, P., Madsen, H. O., Klarlund, M., Svejgaard, A., et al., 2007. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. *Arthritis Rheum,* 56, pp.1446-53.

Pedersen, M., Jacobsen, S., Klarlund, M. & Frisch, M., 2006a. Socioeconomic status and risk of rheumatoid arthritis: a Danish case-control study. *J Rheumatol,* 33, pp.1069-74.

Pedersen, M., Jacobsen, S., Klarlund, M., Pedersen, B. V., Wiik, A., Wohlfahrt, J., et al., 2006b. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. *Arthritis Res Ther,* 8, pp.R133.

Pencina, M. J., D'agostino, R. B., Sr. & Steyerberg, E. W., 2011. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med,* 30, pp.11-21.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W. & Newcomb, P., 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol,* 159, pp.882-90.

Pers, Y. M., Fortunet, C., Constant, E., Lambert, J., Godfrin-Valnet, M., De Jong, A., et al., 2014. Predictors of response and remission in a large cohort of rheumatoid arthritis patients treated with tocilizumab in clinical practice. *Rheumatology (Oxford),* 53, pp.76-84.

Pikwer, M., Bergstrom, U., Nilsson, J. A., Jacobsson, L., Berglund, G. & Turesson, C., 2009. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. *Ann Rheum Dis,* 68, pp.526-30.

Pischon, N., Pischon, T., Kroger, J., Gulmez, E., Kleber, B. M., Bernimoulin, J. P., et al., 2008. Association among rheumatoid arthritis, oral hygiene, and periodontitis. *J Periodontol,* 79, pp.979-86.

Plant, D., Bowes, J., Potter, C., Hyrich, K. L., Morgan, A. W., Wilson, A. G., et al., 2011a. Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci. *Arthritis Rheum,* 63, pp.645-53.

Plant, D., Thomson, W., Lunt, M., Flynn, E., Martin, P., Eyre, S., et al., 2011b. The role of rheumatoid arthritis genetic susceptibility markers in the prediction of erosive disease in patients with early inflammatory polyarthritis: results from the Norfolk Arthritis Register. *Rheumatology (Oxford),* 50, pp.78-84.

Plant, D., Wilson, A. G. & Barton, A., 2014. Genetic and epigenetic predictors of responsiveness to treatment in RA. *Nat Rev Rheumatol,* [Epub ahead of print].

Plant, M. J., Jones, P. W., Saklatvala, J., Ollier, W. E. & Dawes, P. T., 1998. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. *J Rheumatol,* 25, pp.417-26.

Pope, J. E., Bellamy, N. & Stevens, A., 1999. The lack of associations between rheumatoid arthritis and both nulliparity and infertility. *Semin Arthritis Rheum,* 28, pp.342-50.

Posthumus, M. D., Limburg, P. C., Westra, J., Cats, H. A., Stewart, R. E., Van Leeuwen, M. A., et al., 1999. Serum levels of matrix metalloproteinase-3 in relation to the development of radiological damage in patients with early rheumatoid arthritis. *Rheumatology (Oxford),* 38, pp.1081-7.

Posthumus, M. D., Limburg, P. C., Westra, J., Van Leeuwen, M. A. & Van Rijswijk, M. H., 2002. Serum matrix metalloproteinase 3 levels during treatment with sulfasalazine or combination of methotrexate and sulfasalazine in patients with early rheumatoid arthritis. *J Rheumatol,* 29, pp.883-9.

Potter, C., Hyrich, K. L., Tracey, A., Lunt, M., Plant, D., Symmons, D. P. M., et al., 2009. Association of rheumatoid factor and anti-cyclic citrullinated peptide positivity, but not carriage of shared epitope or PTPN22 susceptibility variants, with anti-tumour necrosis factor response in rheumatoid arthritis. *Ann Rheum Dis,* 68, pp.69-74.

Prevoo, M. L., Van 'T Hof, M. A., Kuper, H. H., Van Leeuwen, M. A., Van De Putte, L. B. & Van Riel, P. L., 1995. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum,* 38, pp.44-8.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet,* 38, pp.904-9.

Prince, M. J., 1996. Predicting the onset of Alzheimer's disease using Bayes' theorem. *Am J Epidemiol,* 143, pp.301-8.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al., 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics,* 26, pp.2336-7.

Purcell, S., Cherny, S. S. & Sham, P. C., 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics,* 19, pp.149-50.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81, pp.559-75.

Quirke, A.-M., Fisher, B. a. C., Kinloch, A. J. & Venables, P. J., 2011. Citrullination of autoantigens: upstream of TNFalpha in the pathogenesis of rheumatoid arthritis. *FEBS Lett,* 585, pp.3681-8.

Quirke, A.-M., Lugli, E. B., Wegner, N., Hamilton, B. C., Charles, P., Chowdhury, M., et al., 2014. Heightened immune response to autocitrullinated Porphyromonas gingivalis peptidylarginine deiminase: a potential mechanism for breaching immunologic tolerance in rheumatoid arthritis. *Ann Rheum Dis,* 73, pp.263-9.

Radovits, B. J., Kievit, W., Fransen, J., Van De Laar, M. A., Jansen, T. L., Van Riel, P. L., et al., 2009. Influence of age on the outcome of antitumour necrosis factor alpha therapy in rheumatoid arthritis. *Ann Rheum Dis,* 68, pp.1470-3.

Rahman, M. U., Buchanan, J., Doyle, M. K., Hsia, E. C., Gathany, T., Parasuraman, S., et al., 2011. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. *Ann Rheum Dis,* 70, pp.1631-40.

Rantapaa-Dahlqvist, S., De Jong, B. a. W., Berglin, E., Hallmans, G., Wadell, G., Stenlund, H., et al., 2003. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum,* 48, pp.2741-9.

Rau, R., Wassenberg, S., Herborn, G., Stucki, G. & Gebler, A., 1998. A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol,* 25, pp.2094-107.

Raychaudhuri, S., Sandor, C., Stahl, E. A., Freudenberg, J., Lee, H.-S., Jia, X., et al., 2012. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet,* 44, pp.291-6.

Raza, K., Saber, T. P., Kvien, T. K., Tak, P. P. & Gerlag, D. M., 2012. Timing the therapeutic window of opportunity in early rheumatoid arthritis: proposal for definitions of disease duration in clinical trials. *Ann Rheum Dis,* 71, pp.1921-3.

Reckner Olsson, A., Skogh, T. & Wingren, G., 2001. Comorbidity and lifestyle, reproductive factors, and environmental exposures associated with rheumatoid arthritis. *Ann Rheum Dis,* 60, pp.934-9.

Reinius, L. E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.-E., Greco, D., et al., 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE,* 7, pp.e41361.

Reynisdottir, G., Karimi, R., Joshua, V., Olsen, H., Hensvold, A. H., Harju, A., et al., 2014. Structural changes and antibody enrichment in the lungs are early features of anti-citrullinated protein antibody-positive rheumatoid arthritis. *Arthritis Rheum,* 66, pp.31-9.

Rheumatoid Arthritis Clinical Trial Archive Group, 1995. The effect of age and renal function on the efficacy and toxicity of methotrexate in rheumatoid arthritis. *J Rheumatol,* 22, pp.218-23.

Ripke, S. & Thomas, B. 2011. *Ricopili- a tool for visualizing regions of interest in select GWAS data sets* [Online]. Broad Institute: Broad Institute. Available: http://www.broadinstitute.org/mpg/ricopili/ [Accessed 1st January 2013].

Rodriguez, L. a. G., Tolosa, L. B., Ruigomez, A., Johansson, S. & Wallander, M. A., 2009. Rheumatoid arthritis in UK primary care: incidence and prior morbidity. *Scand J Rheumatol,* 38, pp.173-7.

Rogers, L. M., Riordan, J. D., Swick, B. L., Meyerholz, D. K. & Dupuy, A. J., 2013. Ectopic expression of Zmiz1 induces cutaneous squamous cell malignancies in a mouse model of cancer. *J Invest Dermatol,* 133, pp.1863-9.

Romao, V. C., Canhao, H. & Fonseca, J. E., 2013. Old drugs, old problems: where do we stand in prediction of rheumatoid arthritis responsiveness to methotrexate and other synthetic DMARDs? *BMC Medicine,* 11, pp.17.

Ropes, M. W., Bennett, G. A., Cobb, S., Jacox, R. & Jessar, R. A., 1958. 1958 Revision of diagnostic criteria for rheumatoid arthritis. *Bull Rheum Dis,* 9, pp.175-6.

Rubbert-Roth, A. & Finckh, A., 2009. Treatment options in patients with rheumatoid arthritis failing initial TNF inhibitor therapy: a critical review. *Arthritis Res Ther,* 11 Suppl 1, pp.S1.

Saevarsdottir, S., Wallin, H., Seddighzadeh, M., Ernestam, S., Geborek, P., Petersson, I. F., et al., 2011a. Predictors of response to methotrexate in early DMARD naive rheumatoid arthritis: results from the initial open-label phase of the SWEFOT trial. *Ann Rheum Dis,* 70, pp.469-75.

Saevarsdottir, S., Wedren, S., Seddighzadeh, M., Bengtsson, C., Wesley, A., Lindblad, S., et al., 2011b. Patients with early rheumatoid arthritis who smoke are less likely to respond to treatment with methotrexate and tumor necrosis factor inhibitors: observations from the Epidemiological Investigation of Rheumatoid Arthritis and the Swedish Rheumatology Register cohorts. *Arthritis Rheum,* 63, pp.26-36.

Saleem, B., Mackie, S., Quinn, M., Nizam, S., Hensor, E., Jarrett, S., et al., 2008. Does the use of tumour necrosis factor antagonist therapy in poor prognosis, undifferentiated arthritis prevent progression to rheumatoid arthritis? *Ann Rheum Dis,* 67, pp.1178-80.

Sanmarti, R., Gomez-Centeno, A., Ercilla, G., Larrosa, M., Vinas, O., Vazquez, I., et al., 2007. Prognostic factors of radiographic progression in early rheumatoid arthritis: a two year prospective study after a structured therapeutic strategy using DMARDs and very low doses of glucocorticoids. *Clin Rheumatol,* 26, pp.1111-8.

Schroder, A. E., Greiner, A., Seyfert, C. & Berek, C., 1996. Differentiation of B cells in the nonlymphoid tissue of the synovial membrane of patients with rheumatoid arthritis. *Proc Natl Acad Sci U S A,* 93, pp.221-5.

Scott, D. L. & Choy, E. 2007. *Effect of anakinra (soluble interleukin-1 receptor antagonist) as combination therapy: second UK combination therapy in early rheumatoid arthritis* [Online]. ISRCTN Register. Available: http://www.controlled-trials.com/ISRCTN15819795/15819795 [Accessed 21st April 2014].

Scott, I. C., Lewis, C. M., Cope, A. P. & Steer, S., 2013a. Rheumatoid arthritis severity: Its underlying prognostic factors and how they can be combined to inform treatment decisions. *Int J Clin Rheumatol,* 8, pp.247-263.

Scott, I. C., Seegobin, S. D., Steer, S., Tan, R., Forabosco, P., Hinks, A., et al., 2013b. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. *PLoS Genet,* 9, pp.e1003808.

Scott, I. C., Steer, S., Lewis, C. M. & Cope, A. P., 2011. Precipitating and perpetuating factors of rheumatoid arthritis immunopathology: linking the triad of

genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis. *Best Pract Res Clin Rheumatol,* 25, pp.447-468.

Scott, I. C., Tan, R., Stahl, D., Steer, S., Lewis, C. M. & Cope, A. P., 2013c. The Protective Effect Of Alcohol On Developing Rheumatoid Arthritis: A Systematic Review And Meta-Analysis. *Rheumatology (Oxford),* 52, pp.856-67.

Sebbag, M., Parry, S. L., Brennan, F. M. & Feldmann, M., 1997. Cytokine stimulation of T lymphocytes regulates their capacity to induce monocyte production of tumor necrosis factor-alpha, but not interleukin-10: possible relevance to pathophysiology of rheumatoid arthritis. *Eur J Immunol,* 27, pp.624-32.

Seegobin, S. D., Ma, M. H., Dahanayake, C., Cope, A. P., Scott, D. L., Lewis, C. M., et al., 2014. ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of a randomized controlled trial. *Arthritis Res Ther,* 16, pp.R13.

Shah, T. S., Liu, J. Z., Floyd, J. A., Morris, J. A., Wirth, N., Barrett, J. C., et al., 2012. optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics,* 28, pp.1598-603.

Silman, A., Kay, A. & Brennan, P., 1992. Timing of pregnancy in relation to the onset of rheumatoid arthritis. *Arthritis Rheum,* 35, pp.152-5.

Singh, J. A. & Cameron, D. R., 2012. Summary of AHRQ's comparative effectiveness review of drug therapy for rheumatoid arthritis (RA) in adults--an update. *J Manag Care Pharm,* 18, pp.S1-18.

Singh, J. A., Christensen, R., Wells, G. A., Suarez-Almazor, M. E., Buchbinder, R., Lopez-Olivo, M. A., et al., 2010. Biologics for rheumatoid arthritis: an overview of Cochrane reviews. *Sao Paulo Med J,* 128, pp.309-10.

Singh, J. A., Furst, D. E., Bharat, A., Curtis, J. R., Kavanaugh, A. F., Kremer, J. M., et al., 2012. 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res,* 64, pp.625-39.

Singwe-Ngandeu, M., Finckh, A., Bas, S., Tiercy, J. M. & Gabay, C., 2010. Diagnostic value of anti-cyclic citrullinated peptides and association with HLA-DRB1 shared epitope alleles in African rheumatoid arthritis patients. *Arthritis Res Ther,* 12, pp.R36.

Situnayake, R. D. & Mcconkey, B., 1990. Clinical and laboratory effects of prolonged therapy with sulfasalazine, gold or penicillamine: the effects of disease duration on treatment response. *J Rheumatol,* 17, pp.1268-73.

Soler, G., Radford-Weiss, I., Ben-Abdelali, R., Mahlaoui, N., Ponceau, J. F., Macintyre, E. A., et al., 2008. Fusion of ZMIZ1 to ABL1 in a B-cell acute lymphoblastic leukaemia with a t(9;10)(q34;q22.3) translocation. *Leukemia,* 22, pp.1278-80.

Spector, T. D. & Hochberg, M. C., 1990. The protective effect of the oral contraceptive pill on rheumatoid arthritis: an overview of the analytic epidemiological studies using meta-analysis. *J Clin Epidemiol,* 43, pp.1221-30.

Spector, T. D., Roman, E. & Silman, A. J., 1990. The pill, parity, and rheumatoid arthritis. *Arthritis Rheum,* 33, pp.782-9.

Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., et al., 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet,* 42, pp.508-14.

Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., et al., 2012. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet,* 44, pp.483-9.

Stanczyk, J., Ospelt, C. & Gay, S., 2008. Is there a future for small molecule drugs in the treatment of rheumatic diseases? *Curr Opin Rheumatol,* 20, pp.257-62.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., et al., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology,* 21, pp.128-38.

Strand, V., Kimberly, R. & Isaacs, J. D., 2007. Biologic therapies in rheumatology: lessons learned, future directions. *Nature Reviews. Drug Discovery,* 6, pp.75-92.

Strand, V. & Sharp, J. T., 2003. Radiographic data from recent randomized controlled trials in rheumatoid arthritis: what have we learned? *Arthritis Rheum,* 48, pp.21-34.

Stranzl, T., Wolf, J., Leeb, B. F., Smolen, J. S., Pirker, R. & Filipits, M., 2003. Expression of folylpolyglutamyl synthetase predicts poor response to methotrexate therapy in patients with rheumatoid arthritis. *Clin Exp Rheumatol,* 21, pp.27-32.

Sugiyama, D., Nishimura, K., Tamaki, K., Tsuji, G., Nakazawa, T., Morinobu, A., et al., 2010. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis,* 69, pp.70-81.

Symmons, D., Turner, G., Webb, R., Asten, P., Barrett, E., Lunt, M., et al., 2002. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology (Oxford),* 41, pp.793-800.

Symmons, D. P., Bankhead, C. R., Harrison, B. J., Brennan, P., Barrett, E. M., Scott, D. G., et al., 1997. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. *Arthritis Rheum,* 40, pp.1955-61.

Szekeres-Bartho, J., Barakonyi, A., Par, G., Polgar, B., Palkovics, T. & Szereday, L., 2001. Progesterone as an immunomodulatory molecule. *Int Immunopharmacol,* 1, pp.1037-48.

Talmud, P. J., Hingorani, A. D., Cooper, J. A., Marmot, M. G., Brunner, E. J., Kumari, M., et al., 2010. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ,* 340, pp.b4838.

Taylor, P. C., Steuer, A., Gruber, J., Cosgrove, D. O., Blomley, M. J. K., Marsters, P. A., et al., 2004. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. *Arthritis Rheum,* 50, pp.1107-16.

Teare, M. D., Knevel, R., Morgan, M. D., Kleszcz, A., Emery, P., Moore, D. J., et al., 2013. Allele-dose association of the C5orf30 rs26232 variant with joint damage in rheumatoid arthritis. *Arthritis Rheum,* 65, pp.2555-61.

Teitsson, I., Withrington, R. H., Seifert, M. H. & Valdimarsson, H., 1984. Prospective study of early rheumatoid arthritis. I. Prognostic value of IgA rheumatoid factor. *Ann Rheum Dis,* 43, pp.673-8.

Tennessen, J. A., Bigham, A. W., O'connor, T. D., Fu, W., Kenny, E. E., Gravel, S., et al., 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* 337, pp.64-9.

Turesson, C., O'fallon, W. M., Crowson, C. S., Gabriel, S. E. & Matteson, E. L., 2003. Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years. *Ann Rheum Dis,* 62, pp.722-7.

Turner, S. & Cherry, N., 2000. Rheumatoid arthritis in workers exposed to silica in the pottery industry. *Occup Environ Med,* 57, pp.443-7.

United Kingdom Clinical Research Network. 2013. *PRe-clinical EValuation of Novel Targets in RA (PREVeNT RA): A Nationwide Register of First Degree Relatives of Patients with Rheumatoid Arthritis to Evaluate Predictors of the Development of Rheumatoid Arthritis* [Online]. UK Clinical Research Network: Portfolio Database. Available: http://public.ukcrn.org.uk/search/StudyDetail.aspx?StudyID=14059 [Accessed 7th March 2014].

Van De Stadt, L. A., Witte, B. I., Bos, W. H. & Van Schaardenburg, D., 2013. A prediction rule for the development of arthritis in seropositive arthralgia patients. *Ann Rheum Dis,* 72, pp.1920-6.

Van Der Heijde, D. M., Van Riel, P. L., Van Leeuwen, M. A., Van 'T Hof, M. A., Van Rijswijk, M. H. & Van De Putte, L. B., 1991. Older versus younger onset rheumatoid arthritis: results at onset and after 2 years of a prospective followup study of early rheumatoid arthritis. *J Rheumatol,* 18, pp.1285-9.

Van Der Helm-Van Mil, A. H., Detert, J., Le Cessie, S., Filer, A., Bastian, H., Burmester, G. R., et al., 2008. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: moving toward individualized treatment decision-making. *Arthritis Rheum,* 58, pp.2241-7.

Van Der Helm-Van Mil, A. H., Le Cessie, S., Van Dongen, H., Breedveld, F. C., Toes, R. E. & Huizinga, T. W., 2007. A prediction rule for disease outcome in

patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. *Arthritis Rheum,* 56, pp.433-40.

Van Der Helm-Van Mil, A. H. M. & Huizinga, T. W. J., 2008. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. *Arthritis Res Ther,* 10, pp.205.

Van Der Helm-Van Mil, A. H. M., Kern, M., Gregersen, P. K. & Huizinga, T. W. J., 2006. Variation in radiologic joint destruction in rheumatoid arthritis differs between monozygotic and dizygotic twins and pairs of unrelated patients. *Arthritis Rheum,* 54, pp.2028-30.

Van Der Helm-Van Mil, A. H. M., Verpoort, K. N., Breedveld, F. C., Toes, R. E. M. & Huizinga, T. W. J., 2005. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. *Arthritis Res Ther,* 7, pp.R949-58.

Van Der Linden, M. P. M., Feitsma, A. L., Le Cessie, S., Kern, M., Olsson, L. M., Raychaudhuri, S., et al., 2009. Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis. *Arthritis Rheum,* 60, pp.2242-7.

Van Der Woude, D., Houwing-Duistermaat, J. J., Toes, R. E. M., Huizinga, T. W. J., Thomson, W., Worthington, J., et al., 2009. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis & Rheumatism,* 60, pp.916-23.

Van Der Woude, D., Rantapaa-Dahlqvist, S., Ioan-Facsinay, A., Onnekink, C., Schwarte, C. M., Verpoort, K. N., et al., 2010a. Epitope spreading of the anti-citrullinated protein antibody response occurs before disease onset and is associated with the disease course of early arthritis. *Ann Rheum Dis,* 69, pp.1554-61.

Van Der Woude, D., Syversen, S. W., Van Der Voort, E. I., Verpoort, K. N., Goll, G. L., Van Der Linden, M. P., et al., 2010b. The ACPA isotype profile reflects long-term radiographic progression in rheumatoid arthritis. *Ann Rheum Dis,* 69, pp.1110-6.

Van Dongen, H., Van Aken, J., Lard, L. R., Visser, K., Ronday, H. K., Hulsmans, H. M. J., et al., 2007. Efficacy of methotrexate treatment in patients with probable

rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. *Arthritis Rheum,* 56, pp.1424-32.

Van Steenbergen, H. W., Tsonaka, R., Huizinga, T. W., Le Cessie, S. & Van Der Helm-Van Mil, A. M., 2014. Predicting the severity of joint damage in rheumatoid arthritis; the contribution of genetic factors. *Ann Rheum Dis,* [Epub Ahead of Print].

Van Zeben, D., Hazes, J. M., Zwinderman, A. H., Cats, A., Van Der Voort, E. A. & Breedveld, F. C., 1992. Clinical significance of rheumatoid factors in early rheumatoid arthritis: results of a follow up study. *Ann Rheum Dis,* 51, pp.1029-35.

Vastesaeger, N., Xu, S., Aletaha, D., St Clair, E. W. & Smolen, J. S., 2009. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. *Rheumatology (Oxford),* 48, pp.1114-21.

Verpoort, K. N., Van Gaalen, F. A., Van Der Helm-Van Mil, A. H. M., Schreuder, G. M. T., Breedveld, F. C., Huizinga, T. W. J., et al., 2005. Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. *Arthritis Rheum,* 52, pp.3058-62.

Verstappen, S. M. M., Mccoy, M. J., Roberts, C., Dale, N. E., Hassell, A. B., Symmons, D. P. M., et al., 2010. Beneficial effects of a 3-week course of intramuscular glucocorticoid injections in patients with very early inflammatory polyarthritis: results of the STIVEA trial. *Ann Rheum Dis,* 69, pp.503-9.

Viatte, S., Flynn, E., Lunt, M., Barnes, J., Singwe-Ngandeu, M., Bas, S., et al., 2012. Investigation of Caucasian rheumatoid arthritis susceptibility loci in African patients with the same disease. *Arthritis Res Ther,* 14, pp.R239.

Viatte, S., Plant, D., Lunt, M., Fu, B., Flynn, E., Parker, B. J., et al., 2013. Investigation of rheumatoid arthritis genetic susceptibility markers in the early rheumatoid arthritis study further replicates the TRAF1 association with radiological damage. *J Rheumatol,* 40, pp.144-56.

Visser, K., Goekoop-Ruiterman, Y. P. M., De Vries-Bouwstra, J. K., Ronday, H. K., Seys, P. E. H., Kerstens, P. J. S. M., et al., 2010. A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis

receiving different dynamic treatment strategies: post hoc analyses from the BeSt study. *Ann Rheum Dis,* 69, pp.1333-7.

Vliet Vlieland, T. P., Buitenhuis, N. A., Van Zeben, D., Vandenbroucke, J. P., Breedveld, F. C. & Hazes, J. M., 1994. Sociodemographic factors and the outcome of rheumatoid arthritis in young women. *Ann Rheum Dis,* 53, pp.803-6.

Voigt, L. F., Koepsell, T. D., Nelson, J. L., Dugowson, C. E. & Daling, J. R., 1994. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. *Epidemiology,* 5, pp.525-32.

Wagner, C. A., Sokolove, J., Lahey, L. J., Bengtsson, C., Saevarsdottir, S., Alfredsson, L., et al., 2013. Identification of anticitrullinated protein antibody reactivities in a subset of anti-CCP-negative rheumatoid arthritis: association with cigarette smoking and HLA-DRB1 'shared epitope' alleles. *Ann Rheum Dis,* [Epub ahead of print].

Wajant, H., Henkler, F. & Scheurich, P., 2001. The TNF-receptor-associated factor family: scaffold molecules for cytokine receptors, kinases and their regulators. *Cell Signal,* 13, pp.389-400.

Wallenius, M., Skomsvoll, J. F., Irgens, L. M., Salvesen, K. A., Koldingsnes, W., Mikkelsen, K., et al., 2010. Postpartum onset of rheumatoid arthritis and other chronic arthritides: results from a patient register linked to a medical birth registry. *Ann Rheum Dis,* 69, pp.332-6.

Wang, J., Bansal, A. T., Martin, M., Germer, S., Benayed, R., Essioux, L., et al., 2013. Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. *Pharmacogenomics J,* 13, pp.235-41.

Wang, S., Dvorkin, D. & Da, Y., 2012. SNPEVG: a graphical tool for GWAS graphing with mouse clicks. *BMC Bioinformatics,* 13, pp.319.

Wang, Y., Rollins, S. A., Madri, J. A. & Matis, L. A., 1995. Anti-C5 monoclonal antibody therapy prevents collagen-induced arthritis and ameliorates established disease. *Proc Natl Acad Sci U S A,* 92, pp.8955-9.

Wegner, N., Lundberg, K., Kinloch, A., Fisher, B., Malmstrom, V., Feldmann, M., et al., 2010a. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol Rev,* 233, pp.34-54.

Wegner, N., Wait, R., Sroka, A., Eick, S., Nguyen, K.-A., Lundberg, K., et al., 2010b. Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and -enolase: implications for autoimmunity in rheumatoid arthritis. *Arthritis Rheum,* 62, pp.2662-72.

Wells, G. A., Shea, B., O'connell, D., Peterson, J., Welch, V., Losos, M., et al. 2011. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses* [Online]. Ottawa Hospital Research Institute Website. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [Accessed 3rd December 2011].

Wessels, J. A., Van Der Kooij, S. M., Le Cessie, S., Kievit, W., Barerra, P., Allaart, C. F., et al., 2007. A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. *Arthritis Rheum,* 56, pp.1765-75.

Wingrave, S. J. & Kay, C. R., 1978. Reduction in incidence of rheumatoid arthritis associated with oral contraceptives. Royal College of General Practitioners' Oral Contraception Study. *Lancet,* 1, pp.569-71.

Wolff, B., Berger, T., Frese, C., Max, R., Blank, N., Lorenz, H. M., et al., 2013. Oral status in patients with early rheumatoid arthritis: a prospective, case-control study. *Rheumatology (Oxford),* [Epub ahead of print].

Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M., et al., 2008. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet,* 83, pp.445-56.

Yamanaka, H., Matsuda, Y., Tanaka, M., Sendo, W., Nakajima, H., Taniguchi, A., et al., 2000. Serum matrix metalloproteinase 3 as a predictor of the degree of joint destruction during the six months after measurement, in patients with early rheumatoid arthritis. *Arthritis Rheum,* 43, pp.852-8.

Yamanaka, H., Tanaka, Y., Inoue, E., Hoshi, D., Momohara, S., Hanami, K., et al., 2011. Efficacy and tolerability of tocilizumab in rheumatoid arthritis patients seen in daily clinical practice in Japan: results from a retrospective study (REACTION study). *Mod Rheumatol,* 21, pp.122-133.

Yarwood, A., Han, B., Raychaudhuri, S., Bowes, J., Lunt, M., Pappas, D. A., et al., 2013. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. *Ann Rheum Dis,* [Epub ahead of print].

Young-Min, S., Cawston, T., Marshall, N., Coady, D., Christgau, S., Saxne, T., et al., 2007. Biomarkers predict radiographic progression in early rheumatoid arthritis and perform well compared with traditional markers. *Arthritis Rheum,* 56, pp.3236-47.

SUPPORTING PUBLICATIONS

CLINICAL AND SEROLOGICAL PREDICTORS OF REMISSION IN RHEUMATOID ARTHRITIS ARE DEPENDENT ON TREATMENT REGIMES

MHY Ma, IC Scott, C Dahanayake, AP Cope, DL Scott

Academic Department Of Rheumatology
King's College London
King's College London School of Medicine
Weston Education Centre
King's College London
10 Cutcombe Road
London SE5 9RS

Corresponding author
Dr Margaret Har Yin Ma
Department of Rheumatology
King's College London School of Medicine
Weston Education Centre
King's College London
10 Cutcombe Road
London SE5 9RS
Email: margaret.ma@nhs.net
Telephone: 0207 848 5215
Fax: 0207 848 5202

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in The Journal of Rheumatology following peer review. The definitive publisher-authenticated version J Rheumatol 2014;41(7):1298-1303 is available online at: jrheum@jrheum.org

ABSTRACT

Introduction

Early intensive treatment is now the cornerstone for the management of rheumatoid arthritis. In the era of personalised medicine where treatment is becoming more individualised, it is unclear from the current literature whether all RA patients benefit from such intensive therapies equally. This study investigated the benefit of different treatment regimens on remission rates when stratified to clinical and serological factors.

Methods

The CARDERA trial recruited patients with RA with less than 2 years disease duration who had active disease. The trial compared four treatment regimens: methotrexate monotherapy, 2 different double therapy regimens (methotrexate and ciclosporin or methotrexate and prednisolone) and three-drug therapy. Clinical predictors included age, male and tender joint count (TJC) and serological biomarkers included rheumatoid factor (RF) and antibodies to citrullinated protein antigens (ACPA).

Results

Patients who were male, over 50, had $\geq$ 6 TJC, RF-IgM positive or ACPA positive were more likely to achieve remission at 24 months using three-drug therapy compared to monotherapy (OR 2.99, 4.95, 2.71, 2.54 and 3.52 respectively). There were no differences in response to monotherapy and three-drug therapy if patients were female, under 50, had < 6 TJC or seronegative.

Conclusion

Early intensive regimes have become the gold standard in the treatment of early Rheumatoid arthritis. Our study suggests that this intensive approach is only superior to monotherapy in certain subsets of patients. Although these are unlikely to be the only predictors of treatment response, our study brings us a step closer to achieving personalised medicine in RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a heterogeneous disease with diverse outcomes. Early intensive treatment regimens aiming at achieving remission have been shown to reduce disease activity, structural damage and long-term disability (1-7). This approach is now widely adopted as first-line treatment in routine clinical practice both nationally and internationally (8-10). In the era where personalised medicine is becoming a possiblity, treatment of RA patients should be more individualised. It is unclear from the current literature whether all RA patients benefit from such intensive therapies equally.

We have shown previously that age, gender and baseline tender joint counts (TJC) predict remission at 24 months (11). By using these baseline clinical variables, we developed a remission score which predicted the likelihood of achieving remission at 24 months. While the score is relevant to both clinical trial and routine practice settings, their interaction with treatment was not explored.

Serological biomarkers including rheumatoid factor (RF) and antibodies to citrullinated protein antigens (ACPA) play an important role in the diagnosis of RA (12). The presence of these antibodies is associated with radiographic damage, high disease activity and extra-articular manifestations (13-15). There is emerging evidence that serological status can predict treatment response in biological therapies (16, 17), however, evidence in intensive DMARD therapy is limited (18). In this study, we assessed the role of ACPA and RF status as predictors of remission and evaluated whether clinical and serological biomarkers predict remission in response to different DMARD regimens.

PATIENTS AND METHODS

Patients and samples

The CARDERA trial recruited patients with RA with less than 2 years disease duration who had active disease. Details have been published previously (19). The trial compared four treatment regimens: methotrexate monotherapy, 2 different double therapy regimens (methotrexate and ciclosporin or methotrexate and prednisolone) and three-drug therapy (methotrexate, ciclosporin and prednisolone). Serum samples were taken at baseline.

Autoantibody analysis

Serum samples were taken at baseline. RF-IgM was determined using commercially available ELISA kits (Euroimmun) and expressed as relative units per ml (RU/ml). Testing was performed according to the manufacturer's instructions, at a sample dilution of 1:200. The upper limit of the normal range recommended by Euroimmun is 20 RU/ml. Anti-CCP antibodies (IgG) were measured using an ELISA based kit from Axis-Shield which detects autoantibodies towards a synthetic cyclic peptide containing modified arginine residues (CCP2 peptides). Testing was performed according to the manufacturer's instructions, at a sample dilution of 1:100. The cut-off value for anti-CCP antibody positivity was 5 U/mL.

Remission score

The development of the remission score has been published previously (11). In brief, we used the CARDERA RCT to develop a predictive model for 24-month remission. This model was then validated using data from a UK observational cohort (Early RA Network, ERAN). Remission was defined as 28-joint Disease Activity Score < 2.6. Logistic regression models were used to estimate the associations between remission and potential baseline predictors. Multivariate logistic regression analyses showed age, sex, and tender joint count (TJC) were independently associated with 24-month remission. The multivariate remission score developed using the trial data correctly classified 80% of patients. The Remission Score was

= 0.37 + [-0.03 x age] + [1.1 x gender (1 for males and 0 otherwise] + [-0.07 x Baseline 28TJC]. By combining data from the trial and ERAN, we also developed a simplified remission score that showed that younger men (<50) with a TJC of 5 or lower were most likely to achieve 24-month remission. The effect of treatment was not considered in this paper as treatment differed considerable between the 2 study groups.

Statistical Analysis

Data were analysed using SPSS v20. Analyses were restricted to those individuals with complete data at 24 months and with available serum samples. Remission was defined as DAS28 <2.6 at 24 months. Individual variables were assessed descriptively as median values and interquartile ranges. Categorical data were analysed using Chi-squared test if the n >10 patients or Fisher's exact if n <10 per group. Multiple testing was adjusted by using bonferroni method.

The Remission Score was = 0.37 + [-0.03 x age] + [1.1 x gender (1 for males and 0 otherwise] + [-0.07 x Baseline 28TJC] (11). A higher value indicates a higher probability that the patient will achieve remission at 24 months. Logistic regression modelling was carried out to assess the ability of the remission score to predict remission at 24 months when stratified into different treatment groups. This was adjusted for treatment centre.

Gender, age and baseline tender joint count were dichotomised using thresholds which were used in our previous study (11): Gender – Male or Female, Age - under 50 or over 50 and TJC - < 6 or ≥ 6. Logistic regression models were used to estimate the associations between treatment regimens and point remission at 24 months when stratified by these clinical predictors and serological biomarkers. The effects of treatment on remission rates were first

explored. This showed no difference between double vs monotherapy (OR 0.852 95% CI 0.435 – 1.67, p = ns). The effect of three-drug therapy compared to monotherapy was OR 2.22 95% CI 1.11-4.46 (p = 0.025). The models were therefore restricted to monotherapy vs three-drug therapy with adjustment for treatment centre. To explore the interaction between clinical and serological status, serological status models were also adjusted for baseline DAS28, gender and age.

RESULTS

Study Population

In the CARDERA trial 467 patients were randomised; 378 patients had competed dataset after 24 months of follow up. 351 of these patients had baseline serum samples available for analysis, and so analysis was restricted to these patients. Table 1 summarised their baseline characteristics. There was no difference in baseline DAS28 between patients when stratified according to RF-IgM and to ACPA status: mean initial DAS28 (SD) of RF-IgM negative and positive patients were 5.86 (1.27) and 5.73 (1.29) respectively and of ACPA negative and positive patients were 5.84 (1.36) and 5.69 (1.27).

DAS28 Remission Rates At 24 Months

In total, 16/87 patients (18%), 29/180 (16%) and 30/90 (33%) patients achieved remission at 24 months using monotherapy, double therapy and three-drug therapy respectively. There were no differences between serological status and remission rates at 24 months: 10/44 (23%) of RF-IgM negative and 14/88 (16%) ACPA negative patients achieved remission whereas 65/313 (21%) RF-IgM positive and 60/262 (23%) ACPA positive achieved remission (chi-squared $p > 0.05$).

The Remission Score And Clinical Predictors Of Remission By Treatment Group

The mean (SD) Remission Score was -1.7 (0.84). The Remission Score predicted treatment response in monotherapy, double and three-drug therapy (OR 3.07 95% CI 1.35-6.96 p=0.007, OR 1.99 95% CI 1.19, 3.32 p = 0.008 and OR 4.42 95% CI 1.90 – 8.94, p <0.0001 respectively). This was adjusted for treatment centre.

The individual clinical predictors were then dichotomised: Gender – Male or Female, age - < 50 or ≥ 50 and TJC - < 6 or ≥ 6. 245 patients were female, 113 patients were male, 122 were < 50, 236 were ≥ 50, 88 had less than 6 tender joints and 270 had 6 or more tender joints. Figure 1 shows treatment responses when stratified to different clinical predictors. Females achieved low levels of remission across all treatment arms and responded to a similiar extent to mono-, double and three-drug therapy [8/14 (14%), 17/131 13%, 13/57 23% respectively, p >0.05]. Males responded better to three-drug therapy [17/33, 52%] compared to mono [8/31 26%, 12/49 25%]. Patients with lower TJCs responded to a similar extent across all the treatment groups: mono [6/19, 32%], double [12/44, 27%] and three-drug [10/24, 42%, p = ns]. Patients with more than 6 TJCs achieved higher remission rates with three-drug therapy (20/66, 30%) when compared to mono (10/68 15%) and double (17/136 13%). Patients under 50 achieved similar high rates of remission across all the treatment groups: mono (11/32, 34%), double (14/61 23%) and three-drug (11/29 38%) p = ns). Patients over 50 years of age achieved higher remission rates using three-drug therapy (19/61, 31%) when compared to mono (5/55, 9%) and double (15/119, 13%).

Using logistic regression modelling, patients who were male, over 50 or had ≥ 6 TJC were more likely to achieve remission at 24 months using three-drug therapy compared to

monotherapy (OR 2.99, 4.95 and 2.71 respectively, Table 2). There were no differences in response to monotherapy and three-drug therapy if patients were female, under 50 or had less than 6 tender joints (Table 2).

Serological Predictors Of Remission By Treatment Group

When stratified according to different treatment groups, serological status did have an impact on remission rates (Figure 1). In RF-IgM –ve patients, there was no difference in point remission rates between mono, double and three-drug therapies respectively [2/11 (18%), 5/23 (22%) and 3/10 (30%) p > 0.05]. In RF-IgM +ve patients, fewer patients achieved remission using monotherapy and double therapy (14/76, 18% and 24/157, 15%) when compared to three-drug therapy (27/80, 34%, p = 0.02). In ACPA -ve patients, 5/24 (21%), 4/42 (10%) and 5/22 (23%) achieved remission using mono, double and three-drug therapies respectively (p >0.05*).* In ACPA +ve patients, more patients achieved remission using three-drug therapy (25/67, 37%) than monotherapy (11/63, 17%) and double therapy (24/132, 18%) (p=0.007).

The level of seropositivity was next explored. Patients were stratified into low-positive (< 3 x upper limit of normal) and high-positive (≥ 3 x upper limit of normal) as according to thresholds adopted in the ACR criteria for Rheumatoid Arthritis in 2010 (12). In low-positive RF-IgM, there was no difference between remission rates in the different treatment groups: monotherapy 2/8 (25%), double therapy 0/15 (0%) and three-drug therapy 1/3 (33%, p = ns). In high-positive RF-IgM, more patients achieved remission with three-drug therapy 26/77 (33.8%) than monotherapy 12/68 (17.6%) and double 24/142 (16.9%, p = 0.01). In low-positive ACPA, there was no significant difference in remission rates between the treatment groups: monotherapy 3/5 (60%), double therapy 1/13 (7.7%) and three-drug therapy 2/9

(22%, p = ns). In contrast, in the high-positive ACPA group, more patients achieved remission with three-drug therapy 23/58 (39.7%) when compared to monotherapy 23/76 (13.8%) and double 23/119 (19.3%, $p = 0.001$) groups.

The associations of treatment regimens and remission according to serological status are summarised in Table 2. The benefit of three-drug therapy is only apparent in RF IgM +ve (OR 2.28, 95% CI 1.08-4.85) and ACPA +ve (OR 2.99, 95% CI 1.29-6.97). Their effects size increased when adjusted for clinical factors (DAS28, age and gender) suggesting that the effects of the clinical and serological biomarkers were cumulative (OR 2.54 and 3.52 respectively Table 3).

Serological Status And ACR Core Set Remission Measures

To explore the effects of the individual components of DAS28, the threshold levels for remission according to the ACR core set measures were used (12, 20). At 24 months, in total, 44.7% of patients achieved TJC28 ≤ 1, 22.9% had no swollen joints, 56.2% had ESR ≤ 20 and 23.2% had PGA ≤ 10. There were no differences between monotherapy and three-drug therapy in any of the 4 components at 24 months between RF-IgM positive and negative patients (Table 4). In ACPA +ve patients, more patients achieved TJC28 and SJC28 thresholds of remission in the three-drug therapy group than monotherapy groups at 24 months than ACPA negative patients (Table 4).

DISCUSSION

Early intensive regimes have become the gold standard in the treatment of early Rheumatoid arthritis. Our study suggests that this intensive approach is only superior to monotherapy in certain subsets of patients. Stratifying patients according to gender, age, tender joint counts

(TJC), RF IgM positivity and ACPA positivity can predict those subjects more likely to achieve remission states after 24 months of intensive treatment.

Intensive DMARD therapies are associated with increased drug toxicity (21). A personalised, tailored approach to treatment where each patient receives the appropriate intensity of treatment for as long as needed is the goal of treatment. We have shown previously that female patients of older age, with high TJC were less likely to achieve remission and many other studies have shown similar findings (22-27). However, it may be an over-simplification to suggest that patients with poor prognostic factors will respond to intensive therapies. The current study suggests that males respond better to three-drug therapy compared to monotherapy whereas females respond less well to all treatment regimes. Conversely, patients over 50 and with more than 6 TJC respond better to three-drug therapy than monotherapy but younger patients with less TJC respond well to all treatment regimes.

Prediction matrices using serological status exist to predict risk of rapid radiological progression (RRP) using different DMARD and biological treatment regimens (28). Other studies have shown conflicting results using serological status to predict anti-TNF response (16, 17, 29). However, no model exists for predicting clinical response to intensive DMARD regimes. Our study demonstrates the remission rates of different DMARD regimes are dependent on serological status in early RA patients. This suggests that there may be fundamental differences in the disease of these subsets of patients and treatment regimens should be separated according to serological status.

The main limitation of our study is that it is a post-hoc analysis of an RCT. The findings of

our study will require be further validation in an independent cohort. The treatments used in the RCT - methotrexate, ciclosporin and short-term high dose prednisolone - are not widely used as initial combinations in contemporary RA treatment. Our findings might not be generalisable to all intensive therapies. However, it is a well-recognised combination and many RCTs have demonstrated its efficacy (30-34). Ciclosporin is infrequently used in RA, though there is extensive evidence base for its use, which has been summarised in a Cochrane review by Wells et al. (35). Although it is both effective and relatively safe, other DMARDs like sulfasalazine and hydroxychloroquine are usually given in combination with methotrexate. Thirdly, our study used fixed treatment regimens rather than the treat-to-target approach which is now widely used in early RA management. Our findings suggest further research is needed to assess the benefits and risks of "treat-to-target strategies in ACPA negative disease. Fourthly, we used the DAS28 remission criteria because it is readily achievable in clinical practice. Stricter remission criteria may be preferable in the longer term, such as the ACR/EULAR Boolean remission criteria. Finally, the patients enrolled in CARDERA had more severe early RA than is generally seen in current routine practice.

This study shows a role in a range of conventional clinical and serological biomarkers in predicting treatment responses to combination DMARD therapy. The results suggest that initial combination therapy may only be useful in certain subsets of early RA patients. Although other genetic and laboratory biomarkers are likely to be required to achieve an personalised approach to treatment of RA, our study does challenge the established view that all RA patients should be given combination treatment. Our study favours the more cautious approach in the 2013 EULAR guidance.

Table 1: Baseline Patient Characteristics In 358 Patients With Complete 2 Year Data And Available Serum Samples. IQR = interquartile range, HAQ = Health assessment questionnaire, RF-IgM = Rheumatoid factor IgM isotype, ACPA = antibiodies to citrullinated protein antigens

Clinical Features	Baseline Data
Female n (%)	245 (68%)
Median Age at onset (IQR)	54 (46, 63)
Rheumatoid Nodules n (%)	80 (22%)
Median Baseline DAS28 (IQR)	5.78 (4.88, 6.76)
Median Baseline HAQ (IQR)	1.62 (1.12, 2.03)
Median Larsen Score (IQR)	6.5 (2.3, 16)
RF-IgM positivity n (%)	313 (87%)
ACPA positivity n (%)	258 (72%)

Table 2. Predictive Value Of Achieving Remission At 24 Months Using Three-drug Therapy (Methotrexate, Ciclosporin and Prednisolone) When Compared To Methotrexate Monotherapy Adjusted For Treatment Region

Predictors Of Response	Odds Ratio	95% CI	P Value
Female	1.80	0.68 – 4.78	NS
Male	2.99	1.01 – 8.90	0.049
Over 50	4.95	1.66-14.75	0.004
Under 50	1.09	0.38 – 3.16	NS
≥ 6 TJC	1.56	0.43-5.63	NS
<6 TJC	2.71	1.11-6.60	0.028
RF-IgM Negative	1.49	0.17, 12.46	NS
RF-IgM Positive	2.28	1.08, 4.85	0.032
ACPA Negative	1.03	0.25, 4.30	NS
ACPA Positive	2.99	1.29, 6.97	0.011

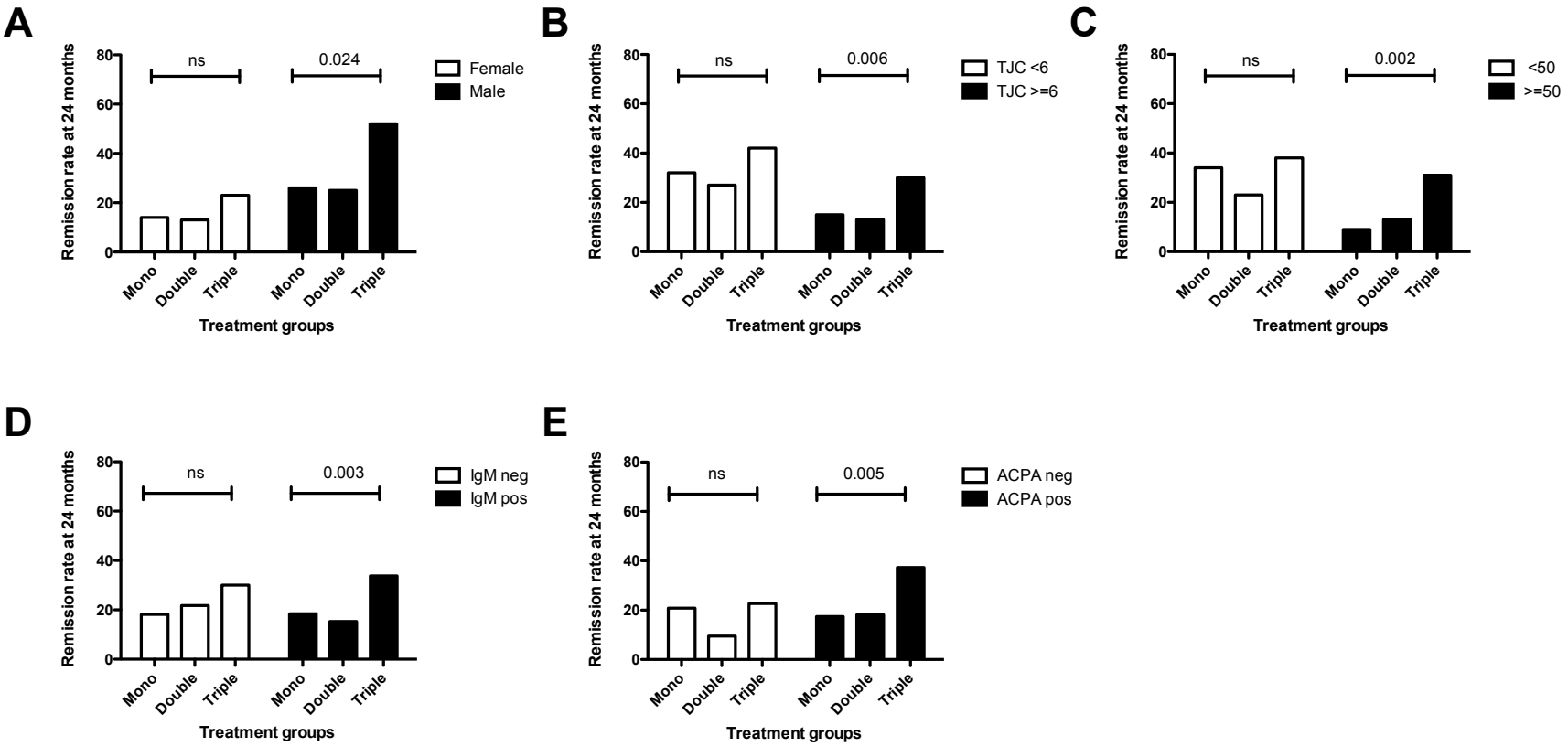
Table 3. The Use Of Serological Status To Predict Remission At 24 Months Using Three-drug Therapy (Methotrexate, Ciclosporin and Prednisolone) Compared To Methotrexate Monotherapy. Adjusted For Treatment Region, Baseline DAS28, Gender And Age

Predictors of response	OR	95% CI	P Value
RF-IgM Negative	1.17	0.58, 23.9	NS
RF-IgM Positive	2.54	1.12, 5.76	0.026
ACPA Negative	0.91	0.19, 4.28	NS
ACPA Positive	3.52	1.37, 9.03	0.009

Table 4. Comparing the effects of Methotrexate monotherapy and three-drug therapy (Methotrexate, Ciclosporin and Prednisolone) in achieving remission scores in the individual components of DAS28 at 24 months when taking into account of serological status. Tender joint count (TJC), Swollen Joint count (SJC), Erythrocyte sedimentation rate (ESR), Patient global assessment (PGA)

Serological status	**Treatment regimes**	**TJC28 at 24 months**		**SJC28 at 24 months**		**ESR at 24 months**		**PGA at 24 months**	
		≤1	*p value*	*<1*	*p value*	*≤20*	*p value*	*≤10*	*p value*
RF IgM Negative	Monotherapy	4/8 (50%)	ns	1/4 (25%)	ns	7/15 (47%)	ns	4/6 (67%)	ns
	Three-drug	4/8 (50%)		3/4 (75%)		8/15 (53%)		2/6 (33%)	
RF IgM Positive	Monotherapy	37/87 (43%)	ns	16/42 (38%)	ns	42/89 (47%)	ns	17/41 (42%)	ns
	Three-drug	50/87 (58%)		26/42 (62%)		47/89 (53%)		24/41 (59%)	
ACPA Negative	Monotherapy	14/25 (56%)	ns	5/10 (50%)	ns	12/27 (44%)	ns	7/13 (54%)	ns
	Three-drug	11/25 (44%)		5/10 (50%)		15/27 (56%)		6/13 (46%)	
ACPA Positive	Monotherapy	27/70 (39%)	0.015	12/36 (33%)	0.033	37/76 (49%)	ns	14/34 (41%)	ns
	Three-drug	43/70 (61%)		24/36 (67%)		39/76 (51%)		20/34 (59%)	

Figure 1: Remission rates at 24 months in different treatment groups according to clinical and serological predictors: (A) Gender, (B) Tender joint count, (C) Age, (D) RF-IgM and (E) ACPA. Multiple testing was adjusted by using bonferroni method. Monotherapy = methotrexate, Double therapy = Methotrexate and Prednsiolone or Methotrexate and Ciclsoporin, Three-drug therapy = Methotrexate, Ciclosporin and Prednisolone



Can we discontinue synthetic disease-modifying anti-rheumatic drugs in rheumatoid arthritis?

I.C. Scott[1,2], G.H. Kingsley[3,4], D.L. Scott[4]

[1]Academic Department of Rheumatology, Centre for Molecular and Cellular Biology of Inflammation, King's College London, London, UK;
[2]Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, UK;
[3]Department of Rheumatology, University Hospital Lewisham, London, UK;
[4]Department of Rheumatology, King's College Hospital, London, UK.

Ian C. Scott, MB ChB, MSc, MRCP
Gabrielle H. Kingsley, MB, ChB, PhD, FRCP
David L. Scott, MD, FRCP

Please address correspondence to:
David L. Scott,
Department of Rheumatology, 3rd Floor, Weston Education Centre,
King's College Hospital,
Cutcombe Road,
London SE5 9RJ, United Kingdom.
E-mail: d.scott1@nhs.net

Received on August 16, 2013; accepted on August 19, 2013.

Clin Exp Rheumatol 2013; 31 (Suppl. 78): S4-S8.

Key words: rheumatoid arthritis, agents anti-rheumatic, autoantibodies

Funding: I.C. Scott receives funding from Arthritis Research UK (grant ref. no. 19739).
G. Kingsley and D.L. Scott are supported by the National Institute for Health Research (NIHR) Programme Grants for Applied Research (http://www.ccf.nihr.ac.uk/PGfAR/Pages/Home.aspx) on "Treatment Intensities and Targets In Rheumatoid Arthritis Therapy: Integrating Patients' And Clinicians' Views – The TITRATE Programme (RP-PG-0610-10066)".

Competing interests: D.L. Scott has received honoraria (<£1000) from Merck Sharp & Dohme Ltd, UCB Pharma and Bristol Myers Squibb within last 3 years; the other co-authors have declared no competing interests.

ABSTRACT

Objective. *When rheumatoid arthritis (RA) patients have achieved sustained good clinical responses can their disease-modifying anti-rheumatic drugs (DMARDs) be reduced or discontinued? This review addresses this question by summarising the clinical evidence about DMARD withdrawal. It includes an assessment of predictive factors for sustained DMARD-free remissions.*

Methods. *We evaluated the evidence for discontinuing DMARDs in stable RA in both randomised controlled trials (RCTs) and observational studies.*

Results. *Six RCTs evaluated DMARD monotherapy withdrawal in 501 RA patients with good clinical responses. Flares occurred in 43/248 (17%) patients who continued DMARD monotherapy and in 117/253 (46%) patients who discontinued DMARDs. Individuals in whom DMARDs were withdrawn were three times more likely to have flares. Restarting DMARDs post-flare was usually successful. Four RCTs evaluated step-down DMARD combinations in comparison to DMARD monotherapy. Patients achieved good clinical responses with combination DMARDs, which were maintained after treatment was tapered to DMARD monotherapy. Four observational studies of tapering or stopping DMARDs in patients with sustained low disease activity states provided supportive evidence for discontinuing DMARDs in some patients. Flares during drug-free remissions were predicted by rheumatoid factor and anti-citrullinated protein antibody status.*

Conclusion. *Drug-free remission is achievable in some RA patients. Discontinuation of DMARDs after patients achieve sustained remissions results in flares in many patients, which can usually be reversed by restarting DMARDs. Step-down DMARD combinations are effective and achieve sustained responses. Further research is required to establish predictors of drug-free remission; these will identify individuals most likely to benefit or experience disease flares after discontinuing DMARDs.*

Introduction

Current rheumatoid arthritis (RA) management emphasises the benefits of early disease-modifying anti-rheumatic drugs (DMARDs), particularly methotrexate, in active disease. Increasing evidence also supports DMARD combinations, which may include glucocorticoids (1, 2). The benefits from using DMARDs extensively must be balanced against patients' wishes to minimise drug use, potential toxicities, and costs of long-term DMARDs. Discontinuing DMARDs when patients achieve sustained low disease activity ameliorates these concerns. It is particularly relevant for DMARD combinations. Some international guidelines recommend reducing DMARDs when patients enter prolonged remissions (3, 4).

The main evidence for discontinuing DMARDs comes from randomised controlled trials (RCTs) in patients with stable RA taking long-term DMARD monotherapy. These RCTs evaluate the impact of stopping treatment on disease activity. Additional evidence comes from RCTs and observational studies in which intensive combination DMARD prescribing follows a step-down approach with combination DMARDs reduced to monotherapy alongside observational studies of stopping DMARDs when patients achieve sustained remission. We summarise these various strands of evidence to provide an overview of the risks and benefits of discontinuing DMARDs.

DMARD retention rates

Strategies for discontinuing DMARDs in good responders must be consid-

ered from the perspective of general retention rates when using DMARDs (5-7). Almost half of patients initiating DMARDs discontinue treatment by 2-3 years. Retention rates differ across DMARDs (Fig. 1). One meta-analysis of 110 studies showed RA patients stay longer on methotrexate than other DMARDs (8). Yazici *et al*. quantified the low risk of discontinuing methotrexate; in 1007 person-years of observation the probability of continuing methotrexate for five years was 79% (9). Low retention rates are commoner in patients receiving combination DMARDs and in those with high disease activity (10).

These low retention rates of patients starting DMARDs mean it is crucial to consider carefully the benefits and risks of discontinuing DMARDs in patients in whom therapy is controlling RA and is not causing adverse effects.

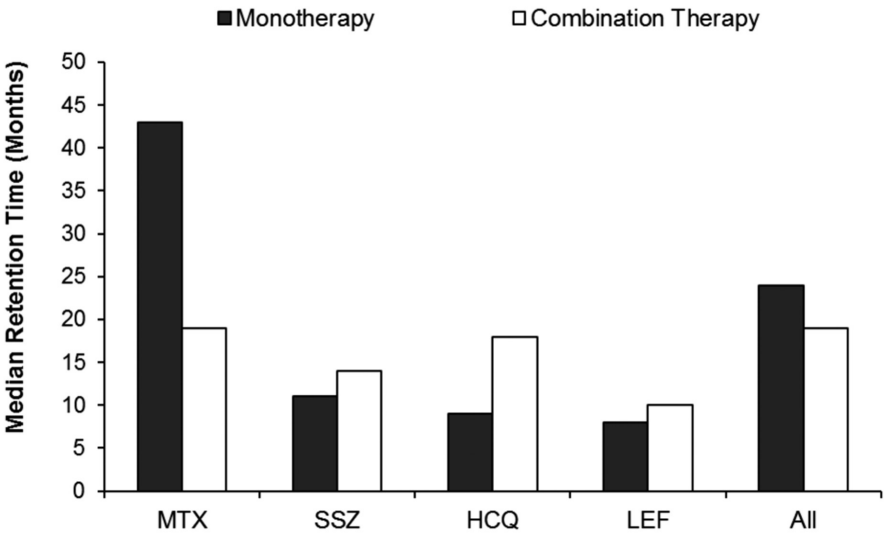


Fig. 1. Retention times on different DMARDs
MTX: methotrexate; SSZ: sulfasalazine; HCQ: hydroxychloroquine; LEF: leflunomide; "All" DMARDs comprise methotrexate, sulfasalazine, hydroxychloroquine, chloroquine, leflunomide, gold, D-penicillamine, azathioprine; Figure adapted using data from the report by Aggarwal *et al*. (5).

Table I. Relative risk of a disease flare in individuals continuing DMARDs compared to those in whom DMARDs were withdrawn.

Study	Year	DMARDs	Patients	Relative risk
Ahern *et al*. (11)	1984	Penicillamine	38	0.13 (0.04, 0.50)
De Silva and Hazleman (12)	1981	Azathioprine	32	0.11 (0.02, 0.73)
Gotzsche *et al*. (13)	1996	Mixed	112	0.25 (0.13, 0.49)
Kremer *et al*. (14)	1987	Methotrexate	10	0.27 (0.07, 1.11)
ten Wolde *et al*. (15)	1996	Mixed	285	0.57 (0.39, 0.84)
Van der Leeden *et al*. (16)	1986	Gold	24	1.18 (0.08, 16.8)
Overall			501	0.31 (0.16, 0.57)

Data from a systematic review and meta-analysis of DMARD withdrawal by O'Mahony *et al*. (17). Pooled relative risks calculated using a random effects model.

Clinical trials examining DMARD withdrawal

Six RCTs published before 2000 evaluated DMARD withdrawal in RA patients in remission or achieving good clinical responses (11-16). The trials, which lasted up to 24 months, enrolled 501 patients. They examined withdrawing a range of DMARD monotherapies including methotrexate, gold, penicillamine and azathioprine. DMARDs were tapered in one RCT (11) and stopped in five RCTs (12-16). The impact of DMARD withdrawal was subsequently evaluated in a meta-analysis by O'Mahony *et al*. (17). It showed that remaining on DMARDs substantially reduced flares (Table I). There were 43/248 (17%) flares in patients staying on DMARDs and 117/253 (46%) flares in patients discontinuing DMARDs. The relative risk of a flare in patients remaining on DMARDs compared to patients in whom DMARDs were stopped was 0.31 (95% confidence interval 0.16 to 0.57; $p<0.001$). Individuals in whom DMARDs were withdrawn were three times more likely to suffer flares than individuals in whom DMARDs were continued.

The largest trial by ten Wolde *et al*. (15) lasted one year and enrolled 285 RA patients with good long-term therapeutic responses. Half the patients continued DMARDs; the others received placebos. The end-point was recurrent synovitis due to flares. By 52 weeks flares had occurred in 38% and 22% of patients receiving placebos and DMARDs, respectively. The trends were similar across all DMARDs (Figure 2), though the study was not powered to compare specific drugs. One limitation in this trial is that it involved very few patients receiving methotrexate. There is evidence that methotrexate achieves better long-term benefits (18) and therefore the benefits of remaining on DMARDs may be underestimated from the perspective of current prescribing practice.

A follow-up study (19) assessed DMARD resumption after flares occurring post-treatment discontinuation. It enrolled 51 patients from the ten Wolde *et al*. trial (15). Patients who had flared showed significant improvements in disease activity measures within three months of restarting DMARDs. Initially they had worse disease activity than before treatment was discontinued. However, by 12 months 35% of patients had inactive disease and 43% had mild disease activity. Only 8% of patients were unable to benefit from resumption of their long-term treatment due to inefficacy.

These studies have a number of limitations: they are small, they include DMARDs that are now rarely used, they have defined flares in a variety of ways and they are of variable quality. Although flares could be controlled by restarting DMARDs, the overall benefit of this strategy was uncertain.

Clinical trials examining step-down DMARDs

Three RCTs evaluated tapering combination DMARDs to monotherapy in strategies based on step-down intensive combination DMARD therapy in early RA. The first step-down RCT was the COBRA early RA trial (20). Its intensive treatment comprised high-dose reducing prednisolone for 28 weeks, low-dose methotrexate for 40 weeks with sulfasalazine as maintenance therapy. Controls received sulfasalazine monotherapy. Both disease activity and erosive progression were better controlled by combination DMARDs. Subsequent follow-up in routine practice settings over 4-5 years showed that the benefits of intensive initial treatment on radiological progression were maintained after tapering (21).

The FIN-RACo trial also assessed step-down treatment (22). It evaluated combination therapy with sulphasalazine, methotrexate, hydroxychloroquine and prednisolone. Treatment was tapered in patients achieving remission during the first year; prednisolone and methotrexate were discontinued. Controls received monotherapy with sulfasalazine followed by methotrexate for patients with adverse effects or non-responders. More patients had good clinical responses and achieved remission with intensive treatment. The radiological benefits were maintained long-term with a subsequent 11-year follow-up report showing less radiologic damage in patients receiving initial combination DMARDs compared to those receiving monotherapy. Mean Larsen score changes over 11 years in the combination and monotherapy groups were 17 (95% CI 12 to 26) and 27 (95% CI 22 to 33), respectively (p=0.037) (2).

Marchesoni *et al*. (23) evaluated maintenance therapy with cyclosporine and methotrexate after 6 months combination treatment with both drugs in 57 early, non-erosive RA patients. Stepping down to single agent maintenance therapy was successful only with methotrexate.

The BeSt study compared four different treatment strategies in 508 patients with recent-onset RA. These comprised DMARD monotherapy, step-up DMARD combinations, step-down DMARD combinations (based on the COBRA regimen) and methotrexate combined with infliximab (1). When patients achieved remission DMARDs were tapered and stopped. Five-year follow-up data evaluated the frequency and impact of DMARD tapering (24). During this period, 23% of patients had drug-free remissions. Subsequently, 46% restarted treatment for increasing disease activity and 51% had drug-free remissions. The frequencies of drug-free remissions were similar across initial treatment groups (Fig. 3). The evidence suggests sustained drug-free remission is uncommon and tapering DMARDs in patients in remission has questionable benefits.

Step-down DMARDs have been evaluated only in a single RCT in established RA. Clegg *et al*. (25) examined if hydroxychloroquine monotherapy extended the benefits of combination therapy with hydroxychloroquine and methotrexate. Patients received open-label combinations of hydroxychloroquine and methotrexate for 24 weeks followed by a double blind period evaluating either methotrexate or hydroxychloroquine as maintenance therapy for 36 weeks. Combination therapy responders were randomised into 3 groups: hydroxychloroquine with methotrexate for flares (40 patients); hydroxychloroquine monotherapy (41 patients); placebo with methotrexate as needed for flares (40 patients). After methotrexate withdrawal, hydroxychloroquine maintenance delayed flare onset (p=0.023). Whilst supporting initial combination therapy, followed by hydroxychloroquine maintenance treatment, this trial did not evaluate methotrexate maintenance therapy, which might be more effective and

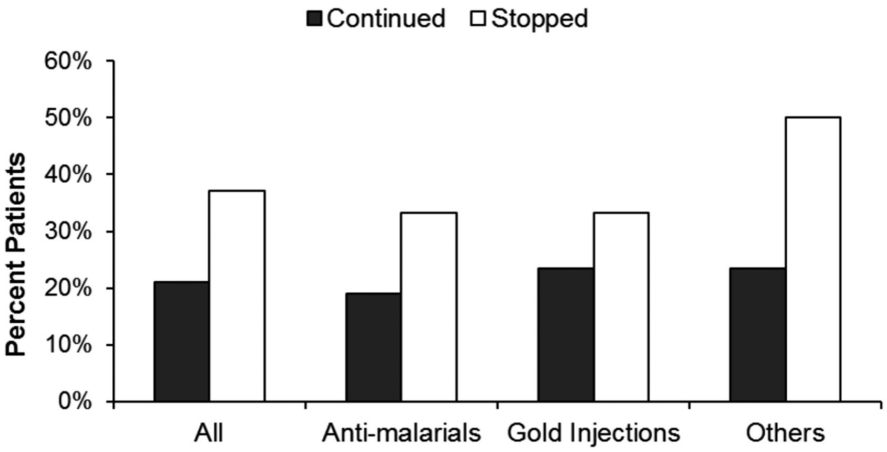


Fig. 2. Cumulative frequency of flares in a trial of DMARD withdrawal. Figure adapted using data from the report by ten Wolde *et al*. (15).

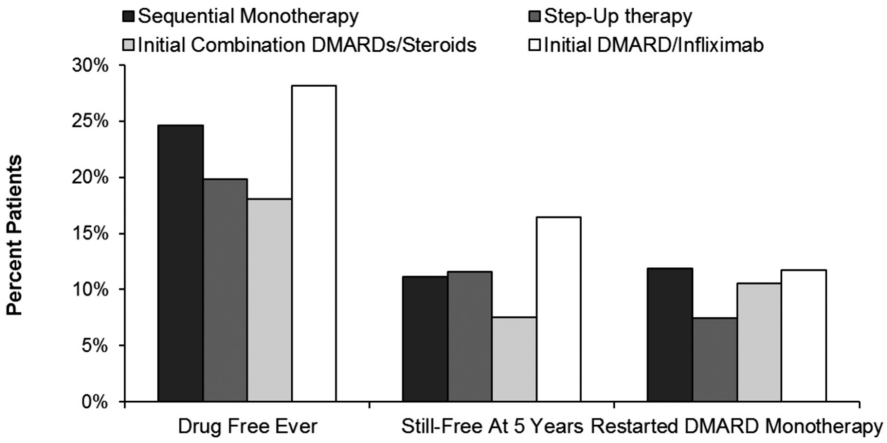


Fig. 3. Drug-free remission rates during five years of follow-up of the BeST early RA study. Figure adapted using data from the report by Klarenbeek *et al*. (24).

when treatment has been stabilised results in low levels of adverse effects.
Overall these RCTs in early and established RA show that step-down combination therapy is effective and has sustained benefits. To reduce subsequent flares at least one anchor DMARD should be retained. The optimal maintenance DMARD regimen was not defined in these RCTs.

Observational studies examining DMARD withdrawal

Frequency reduction

Two very small historical case series examined reducing the frequency of DMARD administration. Reducing methotrexate from weekly to fortnightly in 15 patients in remission showed 13 patients remained in remission and only two flared (26). Reducing penicillamine over 6 months from every day to taking it for one week in four was studied in 14 patients in partial remission on stable treatment (27). Twelve patients had unchanged clinical status over two years and only two flared.

Dose reduction

The 12-month iRAMT trial evaluated reducing methotrexate to a target dose of 5mg/week in patients receiving infliximab (28) in 210 patients. Methotrexate was tapered in the 159 patients with clinical improvements after 22 weeks of infliximab; 92 (58%) subsequently tapered methotrexate without flares. Although it is possible to taper methotrexate when patients have responded to biologics, the overall benefit is uncertain.

Complete withdrawal

The potential of "drug-free" remission as a treatment goal has been reviewed by Goekoop-Ruiterman and Huizinga (29). They noted that in observational studies sustained drug-free remission occurred in 15% of patients in a Dutch Early Arthritis Cohort and 9% of patients in a British cohort (30). The chance of achieving such drug-free remission had not changed over the last two decades. Although stopping DMARDs appears achievable in a small proportion of patients, its constant frequency in different cohorts of patients over time suggests it is the 'natural history' of an RA subset. It most likely represents spontaneous remission without any direct relationship to treatment.
One small 15-year observational study of DMARD withdrawal by Tiippana-Kinnunen *et al*. (31) evaluated DMARD continuity in 70 patients treated since diagnosis with DMARDs following the 'sawtooth' strategy. These patients formed three distinct groups: "continuous DMARDs" (50 patients) receiving continual DMARDs; "discontinued and restarted DMARDs" (9 patients) and "permanently discontinued DMARDs"' (11 patients). In the latter two groups DMARDs were discontinued due to remission lasting at least 12 months or a prolonged symptom-free phase with minor disease activity. Fifteen-year remission rates in these three groups comprised 6%, 0% and 64% respectively. Although DMARDs could be discontinued due to clinical remission or low disease activity states in 29% at 15 years, half of these individuals experienced flares and the overall benefit of stopping treatment is uncertain.

Predicting flare after DMARD withdrawal

Several studies have examined which factors identify individuals attaining sustained drug-free remission on DMARD withdrawal. Van der Woude et al evaluated predictive factors for DMARD-free sustained remission in 454 patients from a Dutch early arthritis clinic and 895 patients from the Early RA Study (ERAS) [30]. Multivariate analyses identified three independent predictors of drug-free remission in both cohorts. These comprised symptom duration, IgM-rheumatoid factor (RF) positivity and presence of the *HLA-DRB1* shared epitope alleles. Of these factors, IgM-RF was by far the strongest predictor with an associated hazard ratio for achieving sustained DMARD-free remission of 0.28 (95% CI 0.16–0.49) in ERAS and 0.19 (95% CI 0.11–0.35) in the Dutch Early Arthritis Clinic; these results show that patients who were IgM-RF positive were far less likely to develop remission that IgM-RF negative patients
Five-year follow-up data from the BeST study also showed that serology predicts drug-free remission (24). Anti-citrullinated protein antibodies (ACPA) positivity was the strongest independent predictor for a flare during drug-free remission (OR 7.5; 95% CI 2.9–19.4). Other predictors of flares included higher mean DAS scores until remission (OR 4.7; 95% CI 1.5–15.2), a lower baseline HAQ (OR 0.41; 95% CI 0.19–0.88) and the use of sulfasalazine as the last DMARD (OR 3.5; 95% CI 1.5–15.2).

Recommendations in international guidelines

After reviewing the available evidence, expert groups have different perspectives about discontinuing DMARDs. There appears to be no overall consensus. UK guidelines from the National Institute for Health and Clinical Excellence (NICE) recommend that if RA is stable, DMARD doses should be cautiously reduced, returning promptly to disease controlling doses if there are any indications of a flare (3). EULAR guidelines are more guarded about DMARD tapering (4). They recommend that in sustained long-term remission cautious titration of synthetic DMARD dose may be considered. By contrast American College of Rheumatology guidelines do not comment on DMARD withdrawal (32).

Conclusions

There is strong evidence from RCTs that treating active RA with step-down DMARD combinations is effective and, in early RA, achieves sustained responses. There is also good evidence that drug-free remission is achievable in a small minority of cases. Many if not most patients who achieve sustained remissions on DMARDs sometimes flare, and the risks of flaring are increased when DMARDs are discontinued, though restarting DMARDs usually reverses these flares. The best current predictors of flares on discontinuing DMARDs are IgM-RF and ACPA-positivity. Further work is required to identify additional predictors of sustained remission on DMARD withdrawal; combining these within

a predictive framework would allow the identification of individuals most likely to benefit from DMARD cessation. Currently, the risks and benefits of stopping DMARD monotherapy in good responders remain uncertain and the evidence for stopping or continuing DMARDs is currently incomplete.

References

1. GOEKOOP-RUITERMAN YPM, DE VRIES-BOUWSTRA JK, ALLAART CF *et al*.: Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum* 2005; 52: 3381-90.
2. RANTALAIHO V, KORPELA M, LAASONEN L *et al*.: Early combination disease-modifying antirheumatic drug therapy and tight disease control improve long-term radiologic outcome in patients with early rheumatoid arthritis: the 11-year results of the Finnish Rheumatoid Arthritis Combination Therapy trial. *Arthritis Res Ther* 2010; 12: R122.
3. DEIGHTON C, O'MAHONY R, TOSH J, TURNER C, RUDOLF M, GUIDELINE DEVELOPMENT GROUP: Management of rheumatoid arthritis: summary of NICE guidance. *BMJ* 2009; 338: b702.
4. SMOLEN JS, LANDEWÉ R, BREEDVELD FC *et al*.: EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis* 2010; 69: 964-75.
5. AGARWAL S, ZAMAN T, HANDA R: Retention rates of disease-modifying anti-rheumatic drugs in patients with rheumatoid arthritis. *Singapore Med J* 2009; 50: 686-92.
6. ALETAHA D, STAMM T, KAPRAL T *et al*.: Survival and effectiveness of leflunomide compared with methotrexate and sulfasalazine in rheumatoid arthritis: a matched observational study. *Ann Rheum Dis* 2003; 62: 944-51.
7. MONTAG K, GINGOLD M, BOERS A, LITTLEJOHN G: Disease-modifying anti-rheumatic drug usage, prescribing patterns and disease activity in rheumatoid arthritis patients in community-based practice. *Intern Med J* 2011; 41: 450-5.
8. MAETZEL A, WONG A, STRAND V, TUGWELL P, WELLS G, BOMBARDIER C: Meta-analysis of treatment termination rates among rheumatoid arthritis patients receiving disease-modifying anti-rheumatic drugs. *Rheumatology* (Oxford) 2000; 39: 975-81.
9. YAZICI Y, SOKKA T, KAUTIAINEN H, SWEARINGEN C, KULMAN I, PINCUS T: Long term safety of methotrexate in routine clinical care: discontinuation is unusual and rarely the result of laboratory abnormalities. *Ann Rheum Dis* 2005; 64: 207-11.
10. MARADIT-KREMERS H, NICOLA PJ, CROWSON CS, O'FALLON WM, GABRIEL SE: Patient, disease, and therapy-related factors that influence discontinuation of disease-modifying antirheumatic drugs: a population-based incidence cohort of patients with rheumatoid arthritis. *J Rheumatol* 2006; 33: 248-55.
11. AHERN MJ, HALL ND, CASE K, MADDISON PJ: D-penicillamine withdrawal in rheumatoid arthritis. *Ann Rheum Dis* 1984; 43: 213-7.
12. DE SILVA M, HAZLEMAN BL: Long-term azathioprine in rheumatoid arthritis: a double-blind study. *Ann Rheum Dis* 1981; 40: 560-3.
13. GOTZSCHE PC, HANSEN M, STOLTENBERG M *et al*.: Randomized, placebo controlled trial of withdrawal of slow-acting antirheumatic drugs and of observer bias in rheumatoid arthritis. *Scand J Rheumatol* 1996; 25: 194-9.
14. KREMER JM, RYNES RI, BARTHOLOMEW LE: Severe flare of rheumatoid arthritis after discontinuation of long-term methotrexate therapy. Double-blind study. *Am J Med* 1987; 82: 781-6.
15. TEN WOLDE S, BREEDVELD FC, HERMANS J *et al*.: Randomised placebo-controlled study of stopping second-line drugs in rheumatoid arthritis. *Lancet* 1996; 347: 347-52.
16. VAN DER LEEDEN H, DIJKMANS BA, HERMANS J, CATS A: A double-blind study on the effect of discontinuation of gold therapy in patients with rheumatoid arthritis. *Clin Rheumatol* 1986; 5: 56-61.
17. O'MAHONY R, RICHARDS A, DEIGHTON C, SCOTT D: Withdrawal of disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a systematic review and meta-analysis. *Ann Rheum Dis* 2010; 69: 1823-6.
18. PINCUS T, MARCUM SB, CALLAHAN LF: Long-term drug therapy for rheumatoid arthritis in seven rheumatology private practices: II. Second line drugs and prednisone. *J Rheumatol* 1992; 19: 1885-94.
19. TEN WOLDE S, HERMANS J, BREEDVELD FC, DIJKMANS BA: Effect of resumption of second line drugs in patients with rheumatoid arthritis that flared up after treatment discontinuation. *Ann Rheum Dis* 1997; 56: 235-9.
20. BOERS M, VERHOEVEN AC, MARKUSSE HM *et al*.: Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. [Erratum appears in *Lancet* 1998; 351: 220]. *Lancet* 1997; 350: 309-18.
21. LANDEWÉ RBM, BOERS M, VERHOEVEN AC *et al*.: COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. *Arthritis Rheum* 2002; 46: 347-56.
22. MOTTONEN T, HANNONEN P, LEIRISALO-REPO M *et al*.: Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FINRACo trial group. *Lancet* 1999; 353: 1568-73.
23. MARCHESONI A, BATTAFARANO N, ARREGHINI M, PANNI B, GALLAZZI M, TOSI S: Radiographic progression in early rheumatoid arthritis: a 12-month randomized controlled study comparing the combination of cyclosporin and methotrexate with methotrexate alone. *Rheumatology* (Oxford) 2003; 42: 1545-9.
24. KLARENBEEK NB, VAN DER KOOIJ SM, GULER-YUKSEL M *et al*.: Discontinuing treatment in patients with rheumatoid arthritis in sustained clinical remission: exploratory analyses from the BeSt study. *Ann Rheum Dis* 2011; 70: 315-9.
25. CLEGG DO, DIETZ F, DUFFY J *et al*.: Safety and efficacy of hydroxychloroquine as maintenance therapy for rheumatoid arthritis after combination therapy with methotrexate and hydroxychloroquine. *J Rheumatol* 1997; 24: 1896-902.
26. TISHLER M, CASPI D, YARON M: Methotrexate treatment of rheumatoid arthritis: is a fortnightly maintenance schedule enough? *Ann Rheum Dis* 1992; 51: 1330-1.
27. DOYLE DV, PERRETT D, FOSTER OJ, ENSOR M, SCOTT DL: The long-term use of D-penicillamine for treating rheumatoid arthritis: is continuous therapy necessary? *Br J Rheumatol* 1993; 32: 614-7.
28. FLEISCHMANN RM, COHEN SB, MORELAND LW *et al*.: Methotrexate dosage reduction in patients with rheumatoid arthritis beginning therapy with infliximab: the Infliximab Rheumatoid Arthritis Methotrexate Tapering (iRAMT) trial. *Curr Med Res Opin* 2005; 21: 1181-90.
29. GOEKOOP-RUITERMAN YPM, HUIZINGA TWJ: Rheumatoid arthritis: can we achieve true drug-free remission in patients with RA? *Nat Rev Rheumatol* 2010; 6: 68-70.
30. VAN DER WOUDE D, YOUNG A, JAYAKUMAR K *et al*.: Prevalence of and predictive factors for sustained disease-modifying antirheumatic drug-free remission in rheumatoid arthritis: results from two large early arthritis cohorts. *Arthritis Rheum* 2009; 60: 2262-71.
31. TIIPPANA-KINNUNEN T, PAIMELA L, KAUTIAINEN H, LAASONEN L, LEIRISALO-REPO M: Can disease-modifying anti-rheumatic drugs be discontinued in long-standing rheumatoid arthritis? A 15-year follow-up. *Scand J Rheumatol* 2010; 39: 12-8.
32. SINGH JA, FURST DE, BHARAT A *et al*.: 2012 update of the 2008 American College of Rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res* 2012; 64: 625-39.

Rheumatoid arthritis severity: its underlying prognostic factors and how they can be combined to inform treatment decisions

Rheumatoid arthritis (RA) is a heterogeneous disease that varies markedly in its severity. There is, therefore, a key research need to develop methods that can predict an individual's likely RA severity at disease onset, which would enable treatment to be tailored accordingly. A number of different prognostic factors for RA severity have been identified. These include environmental (such as smoking and alcohol consumption), genetic (such as the *HLA-DRB1* alleles and polymorphisms in the *IL1* locus), serological (rheumatoid factor and antibodies to citrullinated protein antigens) and biochemical (such as matrix metalloproteinases) factors. In this review, the authors discuss these prognostic factors in detail, outlining the evidence supporting them and focusing on how they have been combined in prognostic modeling to predict the likely severity of an individual's RA phenotype.

KEYWORDS: environmental ■ genotype ■ prognosis ■ rheumatoid arthritis ■ risk factors

Ian C Scott*[1,2], Cathryn M Lewis[2], Andrew P Cope[1] & Sophia Steer[3]

[1]*Academic Department of Rheumatology, Centre for Molecular & Cellular Biology of Inflammation, 1st Floor, New Hunt's House, Guy's Campus, King's College School of Medicine, King's College London, Great Maze Pond, London, SE1 1UL, UK*
[2]*Department of Medical & Molecular Genetics, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London, SE1 9RT, UK*
[3]*Department of Rheumatology, 3rd Floor, Weston Education Centre, King's College Hospital, Cutcombe Road, London, SE5 9RJ, UK*
**Author for correspondence:*
Tel.: +44 20 7188 2601
Fax: +44 20 7188 2585
ian.scott@kcl.ac.uk

Rheumatoid arthritis (RA) is a heterogeneous disease that ranges from a mild, non-erosive form to a severe phenotype characterized by persistent inflammation and rapid radiological progression (RRP). Adopting a uniform, as opposed to a stratified, approach for the management of all RA cases is, therefore, inappropriate and there is a requirement for methods to prospectively establish the likely severity of an individual's RA early in the course of their disease, so that treatments can be tailored accordingly. This 'personalized medicine' approach would limit the development of irreversible articular damage from aggressive RA and prevent exposing individuals with a mild disease course to the potentially toxic effects of multiple drug therapies.

A number of different factors have been shown to associate with RA severity. The evidence underlying many of these is, however, uncertain with environmental and genetic associations often not replicated in independent cohorts. Previous reviews of prognostic factors for RA have either focused on a single factor type, such as genetics [1], described risk factors for a single disease outcome, such as radiological erosions [2], or have not detailed how prognostic factors could be combined to predict RA outcomes [3]. In this review, the authors provide a comprehensive overview of RA severity, outlining the evidence underlying a wide range of prognostic factors – spanning environmental, epidemiological, biochemical, radiological and genetic domains – for multiple RA outcomes with a focus on how they have been combined in prognostic modeling to stratify an individual's risk of severe disease.

The relevance of predicting RA severity: facilitating early treatment in poor prognosis cases

Much evidence exists to support the notion that individuals with RA have better outcomes if treated early and aggressively. The benefits of early combination treatments are demonstrated in several randomized controlled trials (RCTs). These include the BeSt and COBRA studies, with individuals receiving initial combination treatments having significantly better radiographic outcomes compared with those receiving monotherapy or step-up combination therapy, although both RCTs included high-dose steroids or TNF inhibitors in their initial combination regimens, which could explain a significant proportion of their efficacy [4,5]. Evidence also exists that step-up combination disease-modifying antirheumatic drugs (DMARDs) may be as effective as initial combination DMARDs, when used without oral steroids or biologics [6]. Earlier treatment also improves longer-term outcomes: one meta-analysis of 12 observational studies reported a 33% reduction in rates of long-term radiographic progression in patients receiving early versus delayed DMARD therapy [7]. The beneficial effects of prompt aggressive therapies on the natural history of RA have lead to the concepts of 'a window of opportunity' and 'treat to target' in which outcomes are improved provided appropriate treatments are initiated prior

Future Medicine part of fsg

to the end of this window and individuals have their treatment titrated until remission or low disease activity is attained [8,9].

However, despite the benefits of prompt combination therapies, a recent national UK audit of prescribing practices in early RA found that only 50% of 258 rheumatologists surveyed used initial combination treatments in newly diagnosed cases; 81% used sequential monotherapy in at least some patients [10]. The main reasons for this comprised concerns regarding side effects, monitoring requirements and patient acceptability. The capacity to stratify individuals' risks of RA severity at disease onset could facilitate aggressive treatment in the poor prognosis cases that are most likely to benefit from such a management strategy.

Defining RA severity

Although many criteria exist to define remission in RA [11], far fewer criteria have been developed that focus on the opposing end of the disease spectrum, which is defining severe RA. The most widely used criteria in clinical practice is a 28 joint count Disease Activity Score (DAS28) of more than 5.1 [12]. This cross-sectional assessment fails, however, to consider disease severity at more than one time point, disability, erosive disease and the extra-articular impacts of RA. Although several self-reported scales have been developed to assess RA activity, such as the RA Disease Activity Index [13] and the Rapid Assessment of Disease Activity in Rheumatology questionnaire, these do not have thresholds to define severe RA [14]. They also suffer from the same shortcomings as other cross-sectional assessments, and focus on disease activity. The Health Assessment Questionnaire (HAQ) – and more specifically one of its components, the HAQ Disability Index – is commonly used to assess RA severity indirectly through evaluating disability levels. The HAQ Disability Index is a self-reported questionnaire that evaluates functional ability using 20 questions spanning eight categories [15]. Scores of 0–1 are considered to represent mild-to-moderate disability, 1–2 moderate-to-severe disability and 2–3 severe or very severe disability. Separate from clinical criteria, many RCTs and observational studies use radiological damage as indices of RA severity. Two radiological assessments that are commonly used comprise the Sharp/van der Heijde score (SHS) and the Scott modification of the Larsen method, which give scores out of 448 and 250, respectively [16]. Consensus opinion suggests that a change in the SHS of at least 5.0 represents a minimal clinically important difference; the minimal clinically important difference for the modified Larsen score is less clear [17]. A summary of these scoring systems is given in Table 1.

Serological predictors of RA severity

It is increasingly clear that RA is not a single disease entity, but represents a spectrum of clinical syndromes spanning distinct disease subsets [18]. Historically, RA has been stratified according to the presence or absence of rheumatoid factor (RF), termed RF-positive RA (when RF is present) and RF-negative RA (when RF is absent). A more contemporary stratification is by antibodies to citrullinated peptide antigens (ACPA), with ACPA-positive RA characterized by a more aggressive disease course with a greater number of swollen joints and more severe radiological destruction [19]. Interestingly, both ACPA-positive and -negative disease can appear similar at initial presentation [19]; they can also be phenotypically similar at other disease stages.

The role of RF as a predictor of disease severity is well established, with cohorts of RF-positive patients consistently having higher rates of joint damage and extra-articular manifestations. This is particularly true of the IgA RF isotype, which is often reported as having a stronger association with severe disease when compared with IgM and IgG RF [20]. In one longitudinal observational study of 135 women with early RA followed-up for a mean duration of 6 years, while all three RF isotypes were significantly associated with more radiological damage progression and a greater number of swollen joints, IgA RF titers were most strongly correlated with the number of erosions, swollen joint counts (SJCs), the Ritchie index and HAQ scores [21]. Other studies have also shown stronger correlations between IgA RF with radiological erosions [22,23] and extra-articular manifestations [24] in comparison with other RF isotypes.

The prognostic value of ACPA is also well described. In one cohort study of 93 early RA patients identified among Swedish blood donors, the presence of ACPA prior to and at disease onset was significantly associated with radiological outcomes [25]. The baseline and 2-year Larsen scores in cases positive for ACPA pre-disease onset were 8 and 14, respectively; for individuals negative for ACPA pre-disease onset they were 5 and 9. These differences were statistically significant ($p < 0.001$) at both time points. ACPA also predicts longer-term radiological damage. Lindqvist *et al.* demonstrated this point in 183 RA cases followed-up for 10 or more years [26]. In multilinear regression analyses, Larsen

Table 1. Examples of assessment criteria for rheumatoid arthritis severity.

Criteria name	Criteria type	Description	Ref.
DAS28 score	Clinical physician assessment	Composite score involving assessment of the number of swollen and tender joints on a 28 joint count, the ESR and a self-reported VAS of global RA activity A score of >5.1 indicates severe RA activity	[12]
HAQ-DI	Clinical self-reported assessment	Assesses function across eight categories. Following scores represent varying disability levels: 0–1: mild-to-moderate disability 1–2: moderate-to-severe disability 2–3: severe or very severe disability	[15]
Sharp/van der Heijde score	Radiological assessment	Assesses erosions and joint-space narrowing in 44 and 42 joints, respectively, alongside subluxation. The total score ranges from 0 to 448 Higher scores indicate worse disease	[16]
Scott modification of the Larsen method	Radiological assessment	Assesses changes of erosion and joint destruction in the hands, wrists and feet, providing a total score ranging from 0 to 250 Higher scores indicate worse disease	[16]

DAS28: 28 joint count Disease Activity Score; ESR: Erythrocyte sedimentation rate; HAQ-DI: Health Assessment Questionnaire Disability Index; RA: Rheumatoid arthritis; VAS: Visual analog scale.

scores at 10 years were significantly associated with ACPA and C-reactive protein (CRP) levels, which accounted for 32% of the variance in the score.

The prognostic value of antibodies specific for citrullinated peptides in the joint is less certain. Current ACPA assays, such as the anti-CCP2 test, incorporate many peptides derived from proteins absent from the synovial joint; they are, therefore, unlikely to be pathogenic. Although assays specific for citrullinated peptides present within the joint could be more prognostic, current evidence does not support this with a systematic literature review reporting similar associations between antibodies to modified citrullinated vimentin (an intra-articular antigen) and ACPA with radiological progression [27].

Environmental & epidemiological risk factors for RA severity

A variety of environmental and epidemiological factors have been linked with RA severity. These are outlined in Table 2, which also provides examples of which studies have reported this relationship.

Smoking

Cigarette smoking is the dominant environmental risk factor for the development of seropositive RA. A recent systematic review on this topic demonstrated that smoking has a gender-related effect, being associated with RA in men who have smoked at any point, but only being associated with RA in women who have smoked heavily [28]. There is some evidence that cigarette smoking also influences the natural history of RA. In one prospective study of 100 early RA patients followed-up for 24 months, baseline SJC, tender joint count and pain visual analog scale scores were all significantly higher in smokers compared with non-smokers [29]. The SJC at 6 months was also significantly associated with smoking status, with current smoking increasing the number of swollen joints by at least three on average in a regression model after the elimination of non-significant variables. Another observational study of 63 women with advanced RA of an average disease duration of 13.7 years showed that heavy smoking (defined as ≥20 pack-years) was significantly associated with the presence of rheumatoid nodules, higher rates of radiological damage as defined by modified Sharp scores and higher HAQ scores when compared with smokers of <20 pack years or or those who had never smoked [30]. Other studies have, however, failed to demonstrate a clear association between smoking and RA severity with the QUEST-RA study finding no relationship between smoking status and erosions, severe extra-articular disease or DAS28 scores [31]. Any impact of smoking on disease severity probably stems from the fact that it predisposes to the development of ACPA-positive as opposed to ACPA-negative RA in genetically predisposed individuals and, therefore, leads to the onset of a different, more aggressive clinical phenotype [32].

Alcohol consumption

There has been much interest in the role of alcohol consumption as a protective factor against RA development. Recent case–control studies have found lower rates of alcohol consumption in cases compared with controls, implying that it has a protective effect [33]. This relationship has not,

Table 2. Studies evaluating environmental and epidemiological prognostic factors for rheumatoid arthritis severity.

Risk factor	Study (year)	Size	Type	Severity outcome(s)	Main findings	Ref.
Smoking	Masdottir *et al.* (2000)	63 Ca	Cross-sectional	Nodules, modified Sharp score, SJC, HAQ, grip strength	Significant associations between ≥20 pack years and nodules, higher Larsen scores, higher HAQ scores and worse grip strength	[30]
	Manfredsdottir *et al.* (2006)	100 Ca	Longitudinal	Joint counts, pain VAS, CRP, van der Heijde score	Over 24 months current smokers had the highest and those who had never smoked the lowest SJC ($p < 0.001$) and TJC ($p = 0.02$) scores, respectively	[29]
Alcohol	Maxwell *et al.* (2010)	873 Ca	Cross-sectional	Larsen score, DAS28-CRP, modified HAQ, pain VAS	Significant trends for reducing Larsen scores, DAS28-CRP, CRP, modified HAQ and pain VAS with increasing alcohol intake	[34]
	Nissen *et al.* (2010)	2908 Ca	Longitudinal	Ratingen score (radiographic damage), HAQ	Non-significant reduced radiographic progression in drinkers: 1-year mean progression 0.99% (95% CI: 0.89–1.09) in drinkers vs 1.13% (95% CI: 1.01–1.26) in non-drinkers	[35]
OCP	Spector and Hochberg (1990)	1407 Ca 181,081 Co	Meta-analysis	ORs for RA using hospital- or population-derived cases	Pooled OR for studies using hospital cases showed significant protective effect of OCP use on RA development; not observed in studies using population cases	[37]
Periodontitis	Abou-Raya *et al.* (2008)	100 Ca	Cross-sectional	DAS28, HAQ, Larsen score	Periodontitis severity significantly correlated with DAS28 score, ESR and CRP	[40]
	Mercado *et al.* (2001)	65 Ca	Cross-sectional	Joint counts, VAS for physician global/early morning stiffness/pain, ESR/CRP, HAQ	Periodontitis severity significantly associated with higher SJCs, higher HAQ scores and higher CRP/ESR levels	[41]
Gender	Jawaheer *et al.* (2010)	292 Ca	Longitudinal	DAS28, HAQ, pain/fatigue VAS, global health scores, CRP, Sharp scores	Females had worse disease progression reflected by DAS28, physician global and TJC scores	[46]
	Ahlmén *et al.* (2010)	549 Ca	Longitudinal	DAS28, HAQ, SOFI instrument, SHS	Females had significantly higher DAS28 and HAQ scores at all time points	[47]
Social deprivation	McEntegart *et al.* (1997)	814 Ca	Longitudinal	Pain score, articular index, ESR, CRP, HAQ	Cases from deprived areas had significantly higher HAQ scores	[43]
	ERAS Study Group (2000)	869 Ca	Longitudinal	Joint counts, HAQ, pain VAS, grip strength, ESR, erosive radiological changes	Significantly worse HAQ and joint scores, and grip strength in individuals with higher deprivation scores	[44]

Ca: Case; Co: Control; DAS28: 28 joint count Disease Activity Score; ESR: Erythrocyte sedimentation rate; HAQ: Health assessment questionnaire; OCP: Oral contraceptive pill; OR: Odds ratio; RA: Rheumatoid arthritis; SJC: Swollen joint count; SHS: Sharp/van der Heijde score; SOFI: Signals of Functional Impairment; TJC: Tender joint count; VAS: Visual analog scale.

however, been observed in earlier cohort studies. There is also evidence that alcohol intake may associate with a less severe disease course. In one study of 873 erosive RA cases, more frequent alcohol consumption correlated significantly with lower DAS28-CRP, Larsen and modified HAQ scores [34]. These trends are shown in Figure 1. The median DAS28-CRP, Larsen and modified HAQ scores in individuals drinking no alcohol in the month prior to assessment comprised 4.29, 38 and 1.0, respectively; these scores in individuals drinking on more than 10 days in the month prior to assessment comprised 3.72, 27 and 0.63. All of these differences were statistically significant

($p < 0.05$) when evaluated by trend tests across alcohol intake categories. A protective effect of alcohol intake on radiographic progression was also demonstrated in a large Swiss observational study evaluating 2908 RA cases nested within a national database of RA patients [35]. This study evaluated the impact of drinking alcohol on the progression of x-ray damage, scored according to the Ratingen method [36]. It found that in a model adjusting for multiple variables (comprising baseline radiological damage scores, DAS28, HAQ, presence of RF, sex, age, disease duration, tobacco smoking, education level and medications) radiographic damage at 12 months had progressed by an average of 0.99% (95% CI: 0.89–1.09) in drinkers and 1.13% (95% CI: 1.01–1.26) in non-drinkers. Interestingly, as with the beneficial effects of drinking on cardiovascular disease, a J-shaped dose–response effect was seen with occasional and daily alcohol consumers having less radiographic progression at 12 months compared with non-drinkers and heavy drinkers.

Oral contraceptive pill use

Although the oral contraceptive pill (OCP) has often been considered to protect against RA development, a meta-analysis of nine studies evaluating this topic by Spector and Hochberg indicated that OCP use may protect against the progression to a severe RA phenotype as opposed to protecting against disease onset [37]. While an overall protective effect of OCP use on RA risk was observed in case–control studies, when their meta-analysis was subdivided by studies using cases enrolled from hospitals or the community different impacts on disease risk were observed. In case–control studies evaluating hospital-based cases, the odds ratio (OR) for RA in OCP users was 0.49 (95% CI: 0.39–0.63); in those evaluating population-derived cases the OR was 0.95 (95% CI: 0.78–1.16). The authors considered that the most likely explanation for this discrepancy was that rather than preventing RA development, OCP use modified the disease process, maintaining it as a mild or transient disorder.

Periodontitis

Periodontitis, a destructive inflammatory disease of the supporting tissues of the teeth, is prevalent in RA patients [38]. The best characterized causative organism for periodontitis is *Porphyromonas gingivalis*, but there are others, including the *Prevotella* species. *P. gingivalis* is the only known bacterium to express its own functional peptidylarginine deiminase enzyme, the orthologs of the peptidylarginine deiminase family of enzymes responsible for the citrullination of arginine residues in mammals [39]. It has, therefore, been hypothesized that it contributes to ACPA formation in pre-RA individuals. It follows from this that, as with smoking, periodontitis could affect disease severity through promoting ACPA. This relationship was evaluated in a cross-sectional study of 100 patients with active RA, which reported significant correlations between periodontitis severity and DAS28 scores ($p < 0.001$), erythrocyte sedimentation rate (ESR; $p < 0.005$) and high sensitivity CRP levels ($p < 0.003$) [40]. Another small observational study of 65 RA patients found that individuals with moderate-to-severe periodontitis had significantly more swollen joints, higher HAQ scores and higher CRP levels when compared with patients with no or mild periodontitis [41]. Further work is required with large longitudinal studies to better establish this relationship, and to explore the

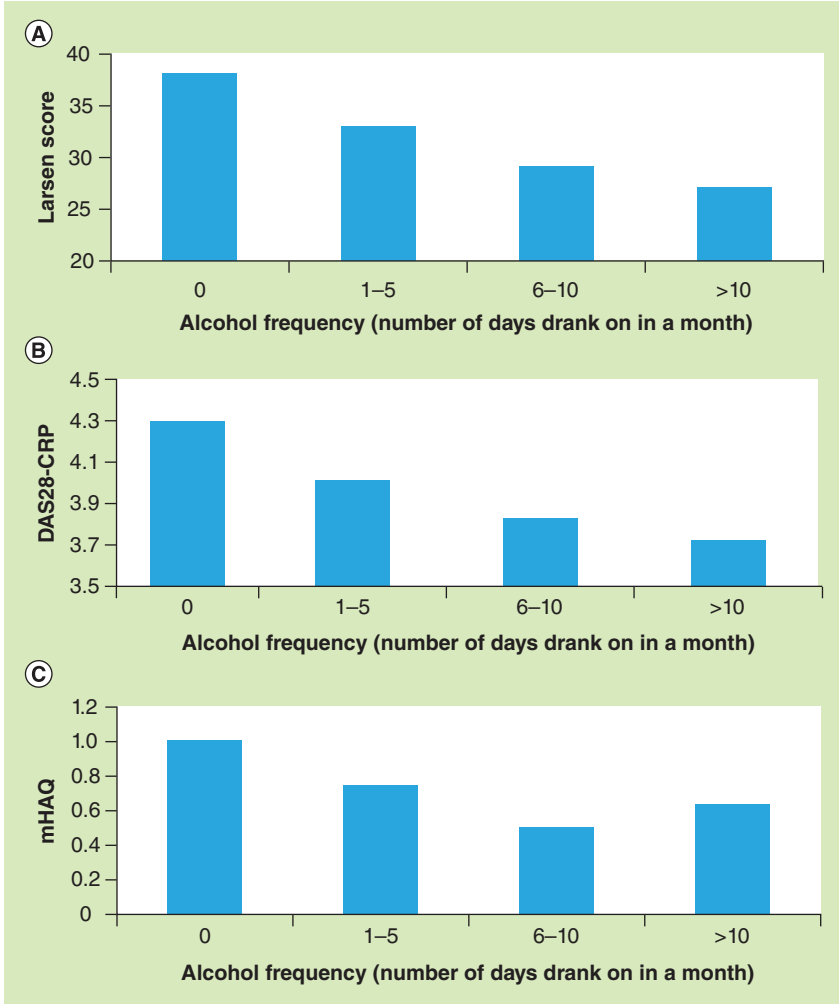


Figure 1. The relationship between alcohol consumption and disease outcomes in a case–control study. p-values for trend tests are all $p < 0.0001$. DAS28: 28 joint count Disease Activity Score; mHAQ: Modified Health Assessment Questionnaire.
Data taken from [34].

impact of aggressive treatment of periodontitis on disease onset and/or severity.

Social deprivation

Several studies have highlighted that individuals from socially deprived areas have poorer disease outcomes [42,43]. This association was evaluated in 869 patients from the Early Rheumatoid Arthritis Study (ERAS), which is a large prospective cohort study of individuals with RA of less than 2 years duration [44]. The authors reported that the Carstairs score (a composite score of male unemployment, social class, overcrowding and car access that represents an index of deprivation) was associated with more severe disease at presentation, as reflected by HAQ and joint scores; this association persisted and remained after 3 years of follow-up. The precise underlying mechanism for this association is unclear; it may represent an association between low socioeconomic status and lifestyle factors such as smoking.

Gender

Gender differences in RA are well described, with the incidence of RA greater amongst women compared with men [45]. There is also evidence that RA outcomes are worse in females. Jawaheer *et al.* found that in a longitudinal prospective study of 225 women and 67 men with early seropositive DMARD-naive RA, women had worse disease progression over 2 years as reflected by DAS28 scores, physician global scores and tender joint counts; this was in spite of similar treatments [46]. Men were also more likely to attain remission. Similarly, the Swedish BARFOT study reported that women had significantly higher DAS28 and HAQ scores compared with males at all time points over a 5-year follow-up period; the authors attributed this DAS28 discrepancy to a higher number of tender joints and general health scores in women compared with men, which suggested that gender differences may exist in pain experiences in RA [47]. Other studies have reported similar female gender influences on RA progression [48].

Evidence for a genetic component to disease severity

In contrast to the identification of genetic susceptibility variants for RA – with 46 loci identified [49] – there is substantially less information on which genetic markers influence RA severity. The dominant reason for this is a lack of adequately sized cohorts containing detailed genotypic and longitudinal disease outcome data. Studies on this topic have been inadequately powered to detect genome-wide significant single nucleotide polymorphisms (SNPs), relying on candidate gene approaches instead to identify loci. While these have had some successes, a candidate gene approach fails to consider the entire genome and important loci may be overlooked. Despite these problems, there is accumulating evidence that genetics play an important role in determining radiological progression in RA. A twin study found that the variance in radiographic joint destruction was highest in unrelated patients, followed by dizygotic and finally monozygotic twins [50]. A more recent study has replicated the association between relatedness and radiological damage in 325 Icelandic patients with RA; this study quantified the heritability of radiological joint destruction to be between 45 and 58% [51].

Genetic risk factors for RA severity

HLA-DRB1 alleles

The *HLA-DRB1* alleles, in particular those encoding the shared epitope (SE), are the best established genetic risk factors for seropositive RA, explaining approximately 36% of the heritability of RA [49]. They also associate with a more severe phenotype [52]. In one case–control study of 309 Caucasian RA patients and 283 controls, heterozygous/homozygous SE carriers used significantly more DMARDs – an indirect marker of disease severity – compared with non-SE carriers [53]. Similarly, Wagner *et al.* found that in a prospective study of 55 early RA cases, those positive for the SE on *HLA-DR4* had an OR for erosive disease of 13.75 ($p = 0.00083$) [54].

More recent studies have employed the classification system for *HLA-DRB1* alleles proposed by du Montcel *et al.* [55]. This broadly divides *HLA-DRB1* alleles into two groups: S alleles and X alleles, which have or do not have the RAA sequence at position 72–74, respectively. Some S alleles – such as S_2 (containing *HLA-DRB1*04:01*) – are associated with an increased disease risk; other S alleles – such as S_1 (containing *HLA-DRB1*13:01*) – are associated with a reduced risk [56]. Using this classification system one observational study of 962 RA cases found that S_2 allele carriage significantly correlated with higher Larsen scores, with the median Larsen score for individuals carrying one and two S_2 copies comprising 29 and 41, respectively [57]. Carriage of S_1 alleles associated with less radiological damage ($p = 0.011$). Similar findings come from a prospective study of 144 French–Caucasian early RA patients in which S_2 allele carriers had greater radiographic damage progression compared with noncarriers ($p = 0.004$); in addition, S_{3D} allele carriers had less

radiographic damage progression compared with noncarriers ($p < 0.0001$) [58]. In both instances significant gene–dose effects were observed.

It therefore appears that, as with disease susceptibility, some *HLA-DRB1* alleles are risk factors for, and some protect against, radiological progression in RA. The association with severity may arise from the fact that carrying SE alleles predisposes individuals to developing ACPA-positive RA [57].

PTPN22

There is limited evidence that *PTPN22* – the dominant non-MHC susceptibility allele – contributes to radiological progression. This allele encodes a lymphoid-specific tyrosine phosphatase, Lyp, which is an important regulator of kinases and signaling intermediates that mediate antigen receptor signal transduction and T-cell activation [59]. It has been suggested that the RA-associated variant represents a gain-of-function mutation that predisposes to autoimmune disease through excessive suppression of T-cell receptor signaling leading to the survival of autoreactive T cells [60], but this remains controversial.

In a cross-sectional study of 964 RA cases, Marinou *et al.* reported a trend towards higher rates of x-ray damage in *PTPN22* minor allele carriers compared with non-carriers. Median modified Larsen scores for individuals with zero, one or two minor allele copies comprised 25.5, 33.0 and 50.0, respectively [61]. This finding was, however, only of borderline statistical significance ($p = 0.04$) and has not been replicated in other cohorts. These include the BRASS in which the adjusted OR (95% CI) for an erosive phenotype in *PTPN22* T allele carriers was 1.14 (0.77–1.71) [62] and the Leiden Early Arthritis Clinic and North American Rheumatoid Arthritis Consortium in which no association was demonstrated between the *PTPN22* susceptibility risk variant and joint destruction rates in RA patients, even when restricting analyses to ACPA-positive RA [63].

IL1B & IL1RN

IL-1 is an important proinflammatory cytokine in RA, as demonstrated by the relative efficacy of anakinra, an IL-1β receptor antagonist [64]. IL-1 induces T-cell activation, promotes lymphocyte and monocyte chemotaxis and facilitates pannus formation. It is, therefore, an ideal candidate gene to examine its role in RA progression. IL-1 comprises three inflammatory mediators, encoded by the *IL1* locus on chromosome 2 [65]. These comprise IL-1α, IL-1β and the IL-1 receptor antagonist (IL-1Ra); all bind to the IL-1 receptor with the initial two mediators stimulating signal transduction and the latter acting as a competitive signaling inhibitor.

Cantagrel *et al.* evaluated the relationship between two polymorphisms in the *IL1B* gene and one polymorphism in the *IL1RN* gene amongst 108 patients with early RA [66]. Although none independently associated with erosion development at 2 years, when *IL1B* exon 5 allele E2 carriage was combined with the presence of SE alleles an increased risk of erosive disease was observed: the OR for erosions was 8.20 (95% CI: 2.59–25.84). This implies that epistasis (gene–gene interactions) contributes to radiological progression. Buchs *et al.* also examined the association between radiological damage and polymorphisms in the *IL1B* (within the promoter region at -511 and in exon 5 at +3954) and *IL1RN* (in exon 2 at position +2018) genes amongst 297 RA cases [67]. They found a significant relationship between destructive RA and the carriage of the rare *IL1B* (+3954) allele 2. The association of the exon 5 +3953 A2 allele with more active RA – defined by higher DAS28 scores and ESR levels – was demonstrated in a smaller study of 93 RA patients [68].

There are a number of studies, however, that show no relationship between *IL1B* loci variants and RA outcomes. In one report of 756 RA patients, three SNPs tagging *IL1B* (rs16944, rs1143623 and rs4848306) and one tagging *IL1A* (rs17561) did not correlate with the presence of rheumatoid nodules, joint replacement need or radiographic progression [69]. Another report found no robust association between 24 SNPs from *IL1A* and *IL1B* and hand radiograph erosions in 712 cases [70].

IL6

IL-6 is another prominent cytokine in RA. It is abundant in the synovial fluid and serum of RA patients; its titers positively correlate with disease activity and joint damage [71]. An association between an *IL6* tagging SNP and radiographic severity was reported by Marinou *et al.* [61]. In this cross-sectional evaluation of 964 RA cases, the SNP, rs1800795, that tags the promoter region of the *IL6* gene (referred to as the '-174' polymorphism) significantly associated with radiological damage in seropositive RA. The modified Larsen scores in ACPA-positive RA risk allele non-carriers, heterozygotes and homozygotes comprised 29, 32 and 41 (trend test p-value = 0.004), respectively. As this finding has not been validated in other cohorts its prognostic relevance is uncertain.

IL10

While many genetic associations in RA are restricted to seropositive disease, one research group identified a polymorphism in *IL10* -592C (tagging SNP, rs1800872) specific for erosive damage in ACPA-negative RA [61]. In this study, ACPA-negative individuals homozygous for the risk allele had more severe radiographic damage compared with non-carriers/heterozygous individuals (pooled due to small numbers); the median modified Larsen score was 6.0 in non-carriers/heterozygotes and 16.0 in homozygotes (p-value for trend = 0.002). This suggests that, as with susceptibility alleles, genetic risks for severity differ serologically. This discrepancy by ACPA status is highlighted in Figure 2, which also demonstrates an *IL6* polymorphism associated with x-ray damage in ACPA-positive, but not ACPA-negative disease. Another *IL10* locus polymorphism was shown to influence the rate of radiological progression in 91 patients in The Netherlands [72]. Although this study did not subdivide its analysis by ACPA status, the presence of the *IL10* -1082GG genotype was associated with significantly greater increases in the Sharp radiographic damage scores at 3 and 6 years when compared with individuals with the -1082AA genotype. Other studies have, however, failed to demonstrate an association between these polymorphisms and radiographic damage in RA [73–75].

IL15

IL-15 is an innate immune system cytokine. It is present in the RA synovium where it plays a functional role, inducing neutrophil activation, granule release from natural killer cells, endothelial cell activation and preventing fibroblast apoptosis [76]. Clinical trials suggest that anti-IL-15 monoclonal antibody treatments may be effective in RA [77], offering further evidence for a pathogenic role. A meta-analysis of 1418 RA patients from four independent data sets evaluated the relationship between polymorphisms in the *IL15* locus and radiographic progression [78]. This involved an initial exploratory analysis in 600 patients from the largest cohort; significant SNPs were subsequently evaluated in the remaining cohorts, with a final combined assessment undertaken. In the initial analysis, five SNPs significantly associated with joint destruction rates. Although not independently replicated in the other data sets (possibly due to limited power in these smaller cohorts) the meta-analysis revealed significant associations for four SNPs. These comprised rs6821171 (protective effect on joint destruction) and rs7667746, rs7665842 and rs4371699 (deteriorative effects). p-values after multiple testing correction comprised rs6821171 ($p = 0.03$), rs7667746 ($p < 0.01$), rs7665842 ($p < 0.01$) and rs4371699 ($p = 0.02$).

TRAF1/C5

TRAF1 encodes an intracellular protein member of the TNF receptor-associated factor family involved in TNF-α signaling [79]; the complement component 5 has been associated with RA in animal models [80]. In the Norfolk Arthritis Register (NOAR) – a primary care-based inception cohort of recent-onset inflammatory polyarthritis patients – two SNPs mapping to the *TRAF1/C5* locus (rs2900180 and rs10760130) were associated with erosions at 5 years; this was independent of ACPA [81]. At 5 years, the ORs for developing erosions in inflammatory polyarthritis after adjusting for ACPA positivity comprised 1.65 (95% CI: 1.13–2.42; $p = 0.01$) for individuals carrying the risk allele for rs2900180 and 1.52 (95% CI: 1.00–2.29; $p = 0.05$) for those carrying the rs10760130 risk allele. The SNP rs2900180 has also been associated with RA patient Larsen scores in ERAS [82]. Although another study of 278 cases reported a significant association between a SNP in this locus (rs10818488), which is in high linkage disequilibrium with rs10760130, and radiological progression [83], this was not reproduced in a meta-analysis of seven data sets (evaluating 2666 RA patients) [84].

CD40

The CD40 protein is expressed on the surface of multiple immune cells; it plays a pivotal role in providing $CD4^+$ T-cell helper activity in immune reactions [85]. The association of the *CD40* locus with RA outcomes was shown in 250 and 393 ACPA-positive RA cases from the Leiden Early Arthritis Clinic cohort and North American Rheumatoid Arthritis Consortium, respectively [86]. In this analysis, the SNP rs4810485 yielded a 1.12-times (95% CI: 1.04–1.21) greater increase in the Sharp score per year in those carrying the risk genotype in the Leiden Early Arthritis Clinic (a significant association remained after correcting for multiple testing). Using a perfect SNP proxy the risk genotype from the Leiden Early Arthritis Clinic cohort also revealed a higher estimated radiological progression rate in the North American Rheumatoid Arthritis Consortium cohort.

Ultrasound imaging & MRI as predictors of RA severity

Advances in imaging technology have lead to an increased use of MRI and, in particular,

musculoskeletal ultrasound scanning (USS) in routine clinical practice. The key advantages that USS has over MRI are that many peripheral joints can be examined multiple times during a consultation with the patient thus improving clinical accuracy, prosthetic joints do not interfere with imaging, and USS is less costly [87]. In early RA, there is evidence that both techniques are able to predict longer-term radiological outcomes.

Ultrasound

Synovial inflammation involves peri-articular vasodilation, synovial proliferation and angiogenesis; this process can be detected by the USS power Doppler (PD) modality [88]. USS, and more specifically PD, assessments have been shown to correlate with radiographic progression in several studies. In 42 early RA patients (with disease duration of less than 12 months) followed-up at 0, 3, 6 and 12 months, time-integrated values of USS PD parameters had stronger correlations with radiographic progression at 1 year ($r = 0.59$; $p < 0.001$) than clinical and laboratory parameters ($r < 0.5$) [87]. In another RCT in which 24 methotrexate-treated RA cases were randomized to either placebo or infliximab, in the placebo arm there were significant positive correlations between both baseline synovial thickness and vascularity as measured by USS and progression in radiographic severity scores at 54 weeks [89].

USS can also play a role in the pre-RA stage by predicting which individuals with an undifferentiated arthritis will develop a persistent disease that may progress to a full RA phenotype. In a study of 50 patients with inflammatory hand symptoms for up to 12 weeks, the presence of a PD score in any joint of at least 2 had a similar predictive value for developing a persistent inflammatory arthritis to that of serology [90]. In this study, the sensitivities/specificities for RF, ACPA and a PD score ≥2 in any joint were reported as 31.6/100.0, 44.7/100.0 and 50.0/100.0, respectively. Similar findings come from a study by Filer *et al.*, who reported that the addition of a 10-joint PD index to the Leiden clinical prediction score for RA development significantly improved the model's predictive capabilities in individuals with very early synovitis (as demonstrated by an area under the curve increase from 0.905 to 0.962; $p < 0.05$) [91].

MRI

Several studies have shown that the presence of MRI-detected bone marrow edema at disease onset predicts joint damage progression years later. In one RCT of 130 early RA patients, baseline MRI bone marrow edema was the only significant predictor (in a multiple linear regression analysis) of radiological progression at the wrist and metacarpophalangeal joints, explaining 41% of the variation in the SHS [92]. Similarly, in a smaller prospective study of 42 RA patients the baseline MRI bone edema score was predictive of the 6-year total Sharp score ($p = 0.01$) [93]. Palosaari *et al.* also demonstrated the predictive value of bone marrow edema on MRI; in 27 early RA patients the baseline MRI bone edema score was the only baseline variable that predicted erosive progression at 24 months in a multivariate model (OR: 4.2; 95% CI: 1.3–13.8) [94].

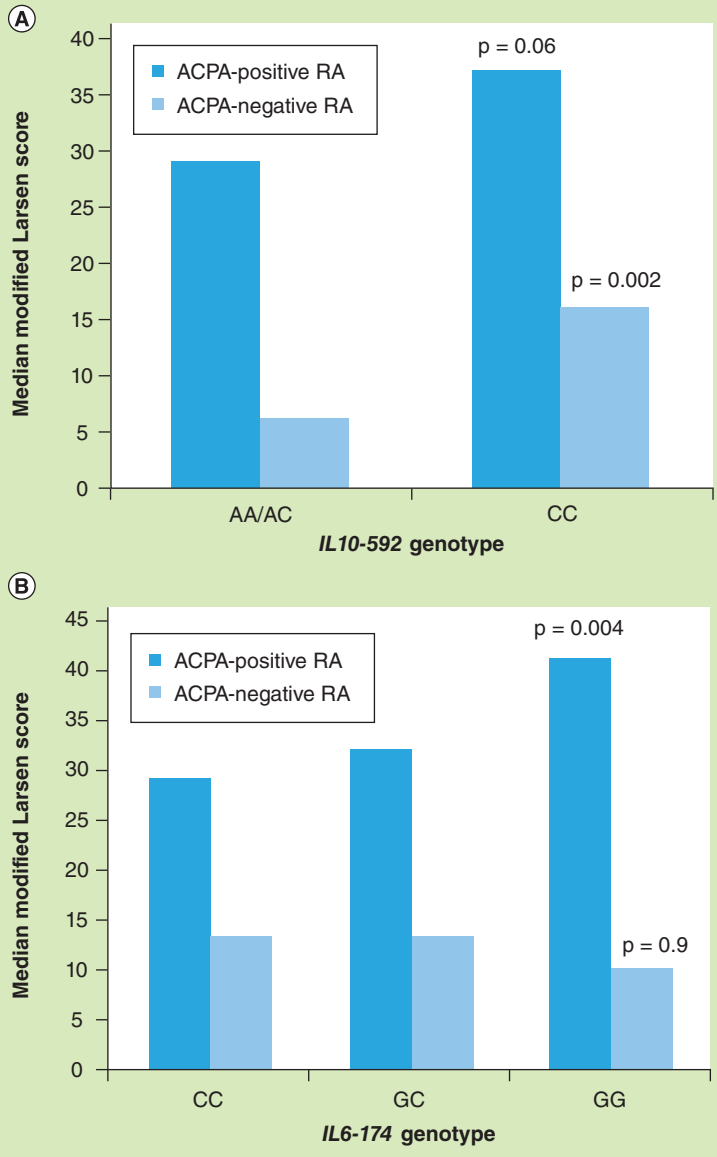


Figure 2. The effects of *IL6* and *IL10* gene polymorphisms on modified Larsen scores when evaluated by ACPA status in a recent case–control study. p-values are for the comparison of median scores across genotype groups. ACPA: Antibodies to citrullinated peptide antigens; RA: Rheumatoid arthritis. Data taken from [61].

Biochemical markers

The best established and most commonly used prognostic biomarkers for RA comprise the acute phase response indices ESR and CRP, both of which correlate with disease severity [26]. A number of other markers have been evaluated for their prognostic implications in RA. One example is the matrix metalloproteinases (MMPs), which are zinc-dependent proteases that regulate extracellular matrix proteolysis and are involved in the cleavage of cytokines, chemokines and their receptors; they are thus considered to play important roles in inflammation [95]. Other examples include the bone turnover marker urinary C-telopeptide of type II collagen (CTX-II), which is an immunoassay that uses antibodies specific for the C-terminal cross-linking telopeptide of type II collagen in the urine [96] and the osteoclast activation markers RANKL and osteoprotegerin (OPG). RANKL is an essential osteoclastogenesis cytokine; OPG is a decoy receptor for RANKL that inhibits osteoclast function by interrupting RANKL's interaction with its receptor [97].

Young-Min *et al.* evaluated the role of several serum biomarkers comprising MMP-1, MMP-13, MMP-3, TIMP-1 and COMP and urinary biomarkers including CTX-II in predicting radiographic progression in 132 early RA patients [98]. They found that although multiple biomarkers including MMP-3, COMP and TIMP-1 correlated significantly with radiographic progression by multivariate analysis, a model consisting of baseline MMP-3 and CTX-II provided the best prediction of radiographic progression at study entry (area under the curve: 0.76; 95% CI: 0.66–0.85). Other research groups have shown MMP-3 to be predictive of radiographic progression in other RA cohorts. In 48 RA patients without radiological damage at presentation, serum MMP-3 levels at study entry significantly correlated with Sharp scores at 6 and 12 months and joint space narrowing at 6, 12 and 24 months [99]. Similarly, in 26 patients with early RA baseline serum MMP-3 levels were significantly associated with Larsen scores at 6 and 12 months after study entry; furthermore, when the relationship between percentage increases in serum MMP-3 in the first 12 months after entry and the percentage increase in Larsen scores in each year were evaluated, a significant correlation was observed between the increase in serum MMP-3 during the first 12 months and the increase in the Larsen score in the subsequent 12–24 months after entry [100].

The role of urinary CTX-II in RA prognostic stratification has also been reproduced in several studies. The association between baseline urinary CTX-I and CTX-II levels and the mean annual progression of joint destruction over a median of 4 years was examined in the COBRA study. In two multivariate logistic regression analyses that included each marker separately due to their high correlation, baseline urinary CTX-I and CTX-II levels both predicted long-term radiologic progression independently of treatment, disease activity and RF status at baseline [101]. In addition, Hashimoto *et al.* reported that in 145 patients with active RA of less than 5 years duration baseline urinary CTX-II levels correlated significantly with radiological progression at week 52 [102].

The prognostic value of the RANKL:OPG ratio (representing osteoclast activation) was also shown in the COBRA study. In a univariate analysis examining the relationships between disease activity measures/bone markers and annual radiographic progression, the baseline RANKL:OPG ratio was the strongest predictor of radiological deterioration [103].

Combining prognostic markers to predict RA severity

Several research groups have attempted to combine information on the aforementioned prognostic factors into models that are capable of identifying individuals at a high risk of radiological progression. Some have used simple clinical parameters and others have integrated these with biomarkers and radiological indices. Genetic markers have rarely been used.

Brennan *et al.* developed one such prediction model for the presence of radiological erosions in the hands and/or feet after 12 months within the NOAR cohort [104]. In this study of 175 patients with early RA, the study population was randomly split into a prediction sample of 105 patients – in which predictor variables for radiological progression were sought – and a validation sample of 70 patients – in which the prediction algorithm was tested. A simple algorithm using a combination of three variables, comprising a positive RF test, swelling of at least two large joints and disease duration of more than 3 months, was best able to predict erosions. This prediction model was able to classify eight risk groups, with a probability of developing erosions that ranged from 0.13 (if all variables were absent) to 0.89 (if all were present). It was able to correctly predict the development of erosive disease in 79% of cases; its negative and positive predictive values were 80 and 76%,

respectively. Its predictive abilities are illustrated in FIGURE 3. This model demonstrates that even simple clinical measurements used in routine practice can be useful in estimating disease progression.

Drossaers-Bakker *et al.* demonstrated that prognostic modeling can be undertaken to predict longer-term disease outcomes at 12 years [105]. This study evaluated 112 female RA patients with symptoms of less than 5 years' duration (median 1 year) at recruitment. It developed prediction models for three different disease outcomes: first, radiographic damage (measured by the SHS method); second, disability (measured by the HAQ); and third, a severe disease course (measured by calculating the area under the curve of all DAS assessments alongside the radiographic disease course). Individuals in the highest tertile of each outcome measure were defined as 'severe' for that outcome and individuals in the lowest tertile were defined as 'mild'. Using a model that contained the baseline parameters of the SJC, RF, the presence of erosions, the Ritchie index, ESR, HAQ and SHS, the accuracy of the model for predicting mild radiographic damage, severe radiographic damage, mild HAQ, severe HAQ and a severe disease course comprised 87, 84, 88, 84 and 83%, respectively. Surprisingly additional information on HLA typing added little to the modeling, improving the correct prediction of radiographic damage by only 3%. This finding highlights the limitations of including current genetic markers, which explain only a minor proportion of the heritability of radiological progression, in prognostic models.

More recently, two research groups have developed matrix risk models for RRP, which are organized into color-coded matrixes similar to that which is widely used in predicting the 10-year risk of fatal cardiovascular disease [106]. One of these matrixes was developed using data from 465 RA patients enrolled to the BeST RCT. As previously described, this study randomized patients to four treatment arms comprising two arms treated with initial monotherapy that could be switched or extended to other DMARDs, a third arm treated with initial combination DMARDs and tapering high-dose corticosteroids and a fourth arm treated with initial methotrexate and infliximab [107]. Patients were treated with an aim of attaining a DAS of ≤2.4. RRP was defined as an increase in the SHS of ≥5 after 12 months. Predictors of RRP were identified by multivariate logistic regression with backward selection. Different models were developed for different treatment groups and included the variables CRP, erosion score and serology (RF and ACPA). The highest risk group was those individuals in the initial monotherapy treatment arm with a CRP ≥35mg/l, erosion score ≥4 and both RF and ACPA positivity; their risk of RRP was 78%. The lowest risk groups were those individuals in the initial combination with prednisolone or

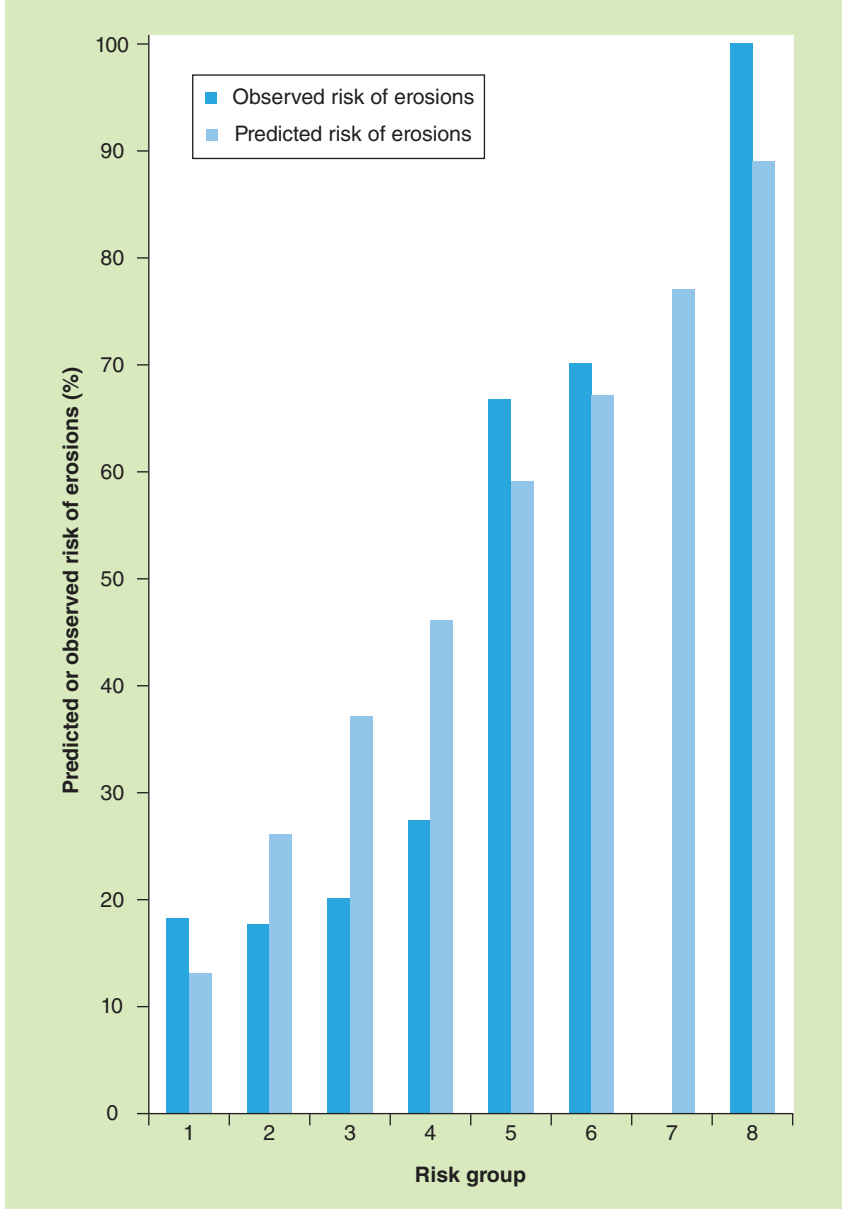


Figure 3. Predicted versus observed risks of developing erosions using a prediction model with three clinical variables in the Norfolk Arthritis Register cohort. The eight risk groups are defined by the presence or absence of the three variables included in the prediction model. The combination of the variables comprising a positive rheumatoid factor, disease duration of ≥3 months and ≥two large joints involved in risk group 1 is negative/no/no; in group 2 is negative/yes/no; in group 3 is negative/no/yes; in group 4 is positive/no/no; in group 5 is negative/yes/yes; in group 6 is positive/yes/no; in group 7 is positive/no/yes; and in group 8 is positive/yes/yes. In risk group 7 there were no individuals with this combination of variables in the Norfolk Arthritis Register validation cohort. Data taken from [104].

infliximab arms with a CRP <10mg/l, erosion score of 0 and negative serology; their risk of RRP was 1%. The area under the curve of the receiver operating curve was 0.81 (95% CI: 0.77–0.86), indicating a moderate ability to correctly classify individuals who will develop RRP.

The other research group to develop a risk matrix for RRP developed prediction models in two different cohorts [106]. The first cohort was the ASPIRE study, which comprised 1049 methotrexate-naive early RA patients randomized to receive methotrexate with or without infliximab; and the second cohort was the ATTRACT trial, which comprised 428 patients with established RA and active disease treated with either methotrexate and infliximab or placebo. They identified risk factors from the early RA cohort (the ASPIRE study) and in order to ensure this combination of risk factors had similar predictive capabilities in a more advanced RA population undergoing similar treatment generated a prediction model in the ATTRACT trial using the same variables. RRP was defined as a change in the modified SHS of ≥5 units/year. Spearman's rank analysis was used to identify baseline risk factors for RRP. Two prediction matrixes were developed, which contained either the ESR or CRP alongside information on the 28 SJC, RF and treatment (monotherapy or combination therapy). The highest risk group was those individuals in the ATTRACT study receiving methotrexate monotherapy with a 28 SJC >17, RF titer >200 U/ml and an ESR >50 mm/h; their risk of RRP was 65%. The lowest risk group was those individuals in the ASPIRE study receiving methotrexate and infliximab with a 28 SJC <10, RF <80 U/ml and an ESR <21 mm/h; their risk of RRP was 2%. Individuals treated with methotrexate monotherapy had higher predicted rates of RRP when compared with those receiving infliximab.

The latter three studies evaluating prognostic modeling that we have described developed and validated their models within the same patient cohorts [105–107]. It is, therefore, expected that they could predict disease outcomes with relative accuracy and their models require further assessment in alternative cohorts to better define their prognostic capabilities.

Conclusion

Research evaluating the prognostic factors for RA has lagged substantially behind that evaluating the underlying risk factors for RA susceptibility. This is particularly true of genetic factors; although 46 RA susceptibility loci of genome-wide significance have been identified, only a handful of risk loci for radiological progression are known. Furthermore, identified loci mainly stem from candidate gene studies of limited sample sizes and have rarely been replicated in independent data sets. Although some evidence suggests an overlap between RA susceptibility and severity loci, for the most part there appears to be little commonality between the two. One key impediment to research in this area is the lack of a consistent definition of what represents 'severe disease' with marked heterogeneity present in the disease severity markers used between studies. We consider that a better classification of severe RA is required to facilitate comparability across studies in this important research field. Another barrier is the lack of large data sets of RA patients with detailed genetic and disease outcome data; this greatly limits the evaluation of genetic predictors of RA outcomes.

Despite these problems prognostic modeling for RA severity has shown some promise, with disease severity prediction models incorporating variables routinely used in clinical practice, such as the ESR and SJCs, showing relative accuracy at identifying those individuals at a high risk of radiological progression. The inclusion of genetic prognostic markers that explain only a minor proportion of the heritability of radiological progression, have added only minor improvements to current prognostic models, highlighting the limited clinical application of current genetic research in this field.

Further work is needed to better define what markers are relevant in predicting RA prognosis. Ideally large longitudinal cohort studies are required that recruit patients at disease onset and capture detailed environmental, genotypic and disease outcome data. Such an approach should identify factors associated with adverse disease outcomes. As many of these factors (such as age, gender and genotypes) will be non-modifiable, the main benefit of their identification lies in their incorporation within prognostic modeling, although it is only factors of large effect sizes that would significantly improve upon existing models. Effective prognostic modeling would facilitate the advent of personalized medicine, allowing treatments to be tailored according to an individual's likelihood of developing severe disease, which is an attractive prospect for both clinicians and patients.

Future perspective

An increased appreciation of the heritability of radiological progression in RA tied in with the

rapid advances in genotyping techniques, such as next-generation sequencing, has placed a key research focus on identifying the genetic variants that influence disease outcomes, such as rapid radiological progression, in RA. We, therefore, envisage that the main area in which this research field will progress is in the identification of novel risk loci for severe RA, which may substantially improve the predictive capabilities of current prognostic models. This could allow the prediction of an individual's risk of severe RA at disease onset, enabling their treatments to be tailored according.

Financial & competing interests disclosure

This manuscript was undertaken as part of an Arthritis Research UK Clinical Research Fellowship (I Scott; Grant ID Number 19739). I Scott and A Cope receive funding for their research into Rheumatoid Arthritis from Arthritis Research UK. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Requirement for prognostic models to predict rheumatoid arthritis severity

- Rheumatoid arthritis (RA) is a heterogeneous disease that varies markedly in its severity. There is, therefore, a requirement to develop methods that can prospectively stratify an individual's risk of severe disease, enabling treatments to be tailored accordingly.

Prognostic factors for RA severity

- Probable environmental and epidemiological prognostic factors for RA severity include smoking, periodontitis, social deprivation and female gender, which are associated with more severe disease, and drinking alcohol and oral contraceptive pill use, which are associated with less severe disease.
- The most reproduced genetic markers for RA severity comprise the *HLA-DRB1* alleles.
- Power Doppler signal on ultrasound and the presence of bone marrow edema on MRI both correlate with subsequent radiological joint damage.
- Biochemical markers such as MMP-3 and urinary C-telopeptide of type II collagen have shown modest capabilities in predicting joint damage.

Current prediction models for RA severity

- Models incorporating clinical prognostic factors have shown some promise in identifying individuals at a high risk of radiological progression.
- Many of these models require validation in separate cohorts of RA patients.

Future work

- A globally accepted definition of 'severe RA' is required to allow comparability across studies examining prognostic factors for RA.
- Further work is needed to better define what markers are relevant in predicting RA prognosis: large, longitudinal cohort studies are required that recruit patients at disease onset and capture detailed environmental, genotypic and disease outcome data.

References

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Marinou I, Maxwell JR, Wilson AG. Genetic influences modulating the radiological severity of rheumatoid arthritis. *Ann. Rheum. Dis.* 69(3), 476–482 (2010).

2 Markatseli TE, Papagoras C, Drosos AA. Prognostic factors for erosive rheumatoid arthritis. *Clin. Exp. Rheumatol.* 28(1), 114–123 (2010).

3 Scott DL. Prognostic factors in early rheumatoid arthritis. *Rheumatology (Oxford)* 39(Suppl. 1), S24–S29 (2000).

4 Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF *et al.* Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum.* 58(Suppl. 2), S126–S135 (2008).

5 Boers M, Verhoeven AC, Markusse HM *et al.* Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet* 350(9074), 309–318 (1997).

6 Saunders SA, Capell HA, Stirling A *et al.* Triple therapy in early active rheumatoid arthritis: a randomized, single-blind, controlled trial comparing step-up and parallel treatment strategies. *Arthritis Rheum.* 58(5), 1310–1317 (2008).

7 Finckh A, Liang MH, van Herckenrode CM, de Pablo P. Long-term impact of early treatment on radiographic progression in rheumatoid arthritis: a meta-analysis. *Arthritis Rheum.* 55(6), 864–872 (2006).

8 Breedveld F. The value of early intervention in RA – a window of opportunity. *Clin. Rheumatol.* 30(Suppl. 1), S33–S39 (2011).

9 Smolen JS, Aletaha D, Bijlsma JWJ *et al.* Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann. Rheum. Dis.* 69(4), 631–637 (2010).

10 Garrood T, Shattles W, Scott DL. Treating early rheumatoid arthritis intensively: current UK practice does not reflect guidelines. *Clin. Rheumatol.* 30(1), 103–106 (2011).

11 Ma MH, Scott IC, Kingsley GH, Scott DL. Remission in early rheumatoid arthritis. *J. Rheumatol.* 37(7), 1444–1453 (2010).

12 Prevoo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* 38(1), 44–48 (1995).

13 Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric

properties and correlation with parameters of disease activity. *Arthritis Rheum.* 38(6), 795–798 (1995).

14 Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (radar) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. *Arthritis Rheum.* 35(2), 156–162 (1992).

15 Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. *Health Qual. Life Outcomes* 1, 20 (2003).

16 Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. *Ann. Rheum. Dis.* 60(9), 817–827 (2001).

17 Bruynesteyn K, van der Heijde D, Boers M *et al.* Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. *Arthritis Rheum.* 46(4), 913–920 (2002).

18 van der Helm-van Mil AH, Huizinga TW. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. *Arthritis Res. Ther.* 10(2), 205 (2008).

19 van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Toes RE, Huizinga TW. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. *Arthritis Res. Ther.* 7(5), R949–R958 (2005).

20 Jonsson T, Valdimarsson H. What about IgA rheumatoid factor in rheumatoid arthritis? *Ann. Rheum. Dis.* 57(1), 63–64 (1998).

21 van Zeben D, Hazes JM, Zwinderman AH, Cats A, van der Voort EA, Breedveld FC. Clinical significance of rheumatoid factors in early rheumatoid arthritis: results of a follow up study. *Ann. Rheum. Dis.* 51(9), 1029–1035 (1992).

22 Teitsson I, Withrington RH, Seifert MH, Valdimarsson H. Prospective study of early rheumatoid arthritis. I. Prognostic value of IgA rheumatoid factor. *Ann. Rheum. Dis.* 43(5), 673–678 (1984).

23 Brik R, Lorber M, Rivkin M, Nahir AM. ELISA determined IgM and IgA rheumatoid factors in seronegative rheumatoid and psoriatic arthritis. *Clin. Exp. Rheumatol.* 8(3), 293–296 (1990).

24 Jonsson T, Arinbjarnarson S, Thorsteinsson J *et al.* Raised IgA rheumatoid factor (RF) but not IgM RF or IgG RF is associated with extra-articular manifestations in rheumatoid arthritis. *Scand. J. Rheumatol.* 24(6), 372–375 (1995).

25 Berglin E, Johansson T, Sundin U *et al.* Radiological outcome in rheumatoid arthritis is predicted by presence of antibodies against cyclic citrullinated peptide before and at disease onset, and by IgA-RF at disease onset. *Ann. Rheum. Dis.* 65(4), 453–458 (2006).

▪▪ **Demonstrates an association between antibodies to citrullinated peptide antigens prior to the onset of rheumatoid arthritis and a greater degree of radiological damage at the time of rheumatoid arthritis diagnosis and 2 years afterwards. This suggests a subclinical process whereby individuals with antibodies to citrullinated peptide antigens have already developed erosions in the absence of a clinically evident inflammatory process.**

26 Lindqvist E, Eberhardt K, Bendtzen K, Heinegard D, Saxne T. Prognostic laboratory markers of joint damage in rheumatoid arthritis. *Ann. Rheum. Dis.* 64(2), 196–201 (2005).

27 Luime JJ, Colin EM, Hazes JMW, Lubberts E. Does anti-mutated citrullinated vimentin have additional value as a serological marker in the diagnostic and prognostic investigation of patients with rheumatoid arthritis? A systematic review. *Ann. Rheum. Dis.* 69(2), 337–344 (2010).

28 Sugiyama D, Nishimura K, Tamaki K *et al.* Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann. Rheum. Dis.* 69(1), 70–81 (2010).

29 Manfredsdottir VF, Vikingsdottir T, Jonsson T *et al.* The effects of tobacco smoking and rheumatoid factor seropositivity on disease activity and joint damage in early rheumatoid arthritis. *Rheumatology (Oxford)* 45(6), 734–740 (2006).

30 Masdottir B, Jónsson T, Manfredsdottir V, Vikingsson A, Brekkan A, Valdimarsson H. Smoking, rheumatoid factor isotypes and severity of rheumatoid arthritis. *Rheumatology (Oxford)* 39(11), 1202–1205 (2000).

31 Naranjo A, Toloza S, Guimaraes Da Silveira I *et al.* Smokers and non smokers with rheumatoid arthritis have similar clinical status: data from the multinational QUEST-RA database. *Clin. Exp. Rheumatol.* 28(6), 820–827 (2010).

32 Klareskog L, Stolt P, Lundberg K *et al.* A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum.* 54(1), 38–46 (2006).

33 Scott IC, Tan R, Stahl D, Steer S, Lewis CM, Cope AP. The protective effect of alcohol on developing rheumatoid arthritis: a systematic review and meta-analysis. *Rheumatology (Oxford)* doi:10.1093/rheumatology/kes376 (2013) (Epub ahead of print).

34 Maxwell JR, Gowers IR, Moore DJ, Wilson AG. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. *Rheumatology (Oxford)* 49(11), 2140–2146 (2010).

▪ **Demonstrates an inverse relationship between drinking alcohol and rheumatoid arthritis severity outcomes.**

35 Nissen MJ, Gabay C, Scherer A, Finckh A, Swiss Clinical Quality Management Project in Rheumatoid A. The effect of alcohol on radiographic progression in rheumatoid arthritis. *Arthritis Rheum.* 62(5), 1265–1272 (2010).

36 Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. *J. Rheumatol.* 25(11), 2094–2107 (1998).

37 Spector TD, Hochberg MC. The protective effect of the oral contraceptive pill on rheumatoid arthritis: an overview of the analytic epidemiological studies using meta-analysis. *J. Clin. Epidemiol.* 43(11), 1221–1230 (1990).

38 de Pablo P, Dietrich T, Mcalindon TE. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. *J. Rheumatol.* 35(1), 70–76 (2008).

39 Mangat P, Wegner N, Venables PJ, Potempa J. Bacterial and human peptidylarginine deiminases: targets for inhibiting the autoimmune response in rheumatoid arthritis? *Arthritis Res. Ther.* 12(3), 209 (2010).

40 Abou-Raya S, Abou-Raya A, Naim A, Abuelkheir H. Rheumatoid arthritis, periodontal disease and coronary artery disease. [Erratum in *Clin. Rheumatol.* 27(4), 551 (2008)]. *Clin. Rheumatol.* 27(4), 421–427 (2008).

41 Mercado FB, Marshall RI, Klestov AC, Bartold PM. Relationship between rheumatoid arthritis and periodontitis. *J. Periodontol.* 72(6), 779–787 (2001).

42 Vliet Vlieland TP, Buitenhuis NA, Van Zeben D, Vandenbroucke JP, Breedveld FC, Hazes JM. Sociodemographic factors and the outcome of rheumatoid arthritis in young women. *Ann. Rheum. Dis.* 53(12), 803–806 (1994).

43 McEntegart A, Morrison E, Capell HA *et al.* Effect of social deprivation on disease severity and outcome in patients with rheumatoid

arthritis. *Ann. Rheum. Dis.* 56(7), 410–413 (1997).

44 Socioeconomic deprivation and rheumatoid disease: what lessons for the health service? ERAS Study Group. Early Rheumatoid Arthritis Study. *Ann. Rheum. Dis.* 59(10), 794–799 (2000).

45 Myasoedova E, Crowson CS, Kremers HM, Therneau TM, Gabriel SE. Is the incidence of rheumatoid arthritis rising? Results from Olmsted County, Minnesota, 1955–2007. *Arthritis Rheum.* 62(6), 1576–1582 (2010).

46 Jawaheer D, Maranian P, Park G, Lahiff M, Amjadi SS, Paulus HE. Disease progression and treatment responses in a prospective DMARD-naive seropositive early rheumatoid arthritis cohort: does gender matter? *J. Rheumatol.* 37(12), 2475–2485 (2010).

▪▪ **Indicates a likely gender difference in rheumatoid arthritis severity, with females having worse disease progression despite receiving similar treatments to males.**

47 Ahlmén M, Svensson B, Albertsson K, Forslind K, Hafstrom I, Group BS. Influence of gender on assessments of disease activity and function in early rheumatoid arthritis in relation to radiographic joint damage. *Ann. Rheum. Dis.* 69(1), 230–233 (2010).

48 Kuiper S, van Gestel AM, Swinkels HL, de Boo TM, da Silva JA, van Riel PL. Influence of sex, age, and menopausal state on the course of early rheumatoid arthritis. *J. Rheumatol.* 28(8), 1809–1816 (2001).

49 Eyre S, Bowes J, Diogo D *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* 44(12), 1336–1340 (2012).

50 van der Helm-van Mil AH, Kern M, Gregersen PK, Huizinga TW. Variation in radiologic joint destruction in rheumatoid arthritis differs between monozygotic and dizygotic twins and pairs of unrelated patients. *Arthritis Rheum.* 54(6), 2028–2030 (2006).

51 Knevel R, Grondal G, Huizinga TW *et al.* Genetic predisposition of the severity of joint destruction in rheumatoid arthritis: a population-based study. *Ann. Rheum. Dis.* 71(5), 707–709 (2012).

▪▪ **Quantifies the heritability of radiological joint destruction in rheumatoid arthritis in a unique Icelandic population with detailed genealogy information. The high heritability rates highlight a key research need to better determine genetic predictors of rheumatoid arthritis severity.**

52 Gonzalez-Gay MA, Garcia-Porrua C, Hajeer AH. Influence of human leukocyte antigen-DRB1 on the susceptibility and severity of rheumatoid arthritis. *Semin. Arthritis Rheum.* 31(6), 355–360 (2002).

53 Meyer JM, Evans TI, Small RE *et al.* *HLA-DRB1* genotype influences risk for and severity of rheumatoid arthritis. *J. Rheumatol.* 26(5), 1024–1034 (1999).

54 Wagner U, Kaltenhauser S, Sauer H *et al.* HLA markers and prediction of clinical course and outcome in rheumatoid arthritis. *Arthritis Rheum.* 40(2), 341–351 (1997).

55 du Montcel ST, Michou L, Petit-Teixeira E *et al.* New classification of *HLA-DRB1* alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum.* 52(4), 1063–1068 (2005).

56 Barnetche T, Constantin A, Cantagrel A, Cambon-Thomsen A, Gourraud PA. New classification of *HLA-DRB1* alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. *Arthritis Res. Ther.* 10(1), R26 (2008).

57 Mewar D, Marinou I, Coote AL *et al.* Association between radiographic severity of rheumatoid arthritis and shared epitope alleles: differing mechanisms of susceptibility and protection. *Ann. Rheum. Dis.* 67(7), 980–983 (2008).

▪ **Demonstrates that some *HLA-DRB1* alleles are associated with increased and some with reduced levels of radiological progression.**

58 Gourraud PA, Boyer JF, Barnetche T *et al.* A new classification of *HLA-DRB1* alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity. *Arthritis Rheum.* 54(2), 593–599 (2006).

59 Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM. Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. *Blood* 93(6), 2013–2024 (1999).

60 Bottini N, Vang T, Cucca F, Mustelin T. Role of *PTPN22* in Type 1 diabetes and other autoimmune diseases. *Semin. Immunol.* 18(4), 207–213 (2006).

61 Marinou I, Healy J, Mewar D *et al.* Association of interleukin-6 and interleukin-10 genotypes with radiographic damage in rheumatoid arthritis is dependent on autoantibody status. *Arthritis Rheum.* 56(8), 2549–2556 (2007).

62 Karlson EW, Chibnik LB, Cui J *et al.* Associations between human leukocyte antigen, *PTPN22*, *CTLA4* genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study. *Ann. Rheum. Dis.* 67(3), 358–363 (2008).

63 Van Nies JAB, Knevel R, Daha N *et al.* The *PTPN22* susceptibility risk variant is not associated with the rate of joint destruction in anti-citrullinated protein antibody-positive rheumatoid arthritis. *Ann. Rheum. Dis.* 69(9), 1730–1731 (2010).

64 Furst DE. Anakinra: review of recombinant human interleukin-I receptor antagonist in the treatment of rheumatoid arthritis. *Clin. Ther.* 26(12), 1960–1975 (2004).

65 Nicklin MJ, Weith A, Duff GW. A physical map of the region encompassing the human interleukin-1α, interleukin-1β, and interleukin-1 receptor antagonist genes. *Genomics* 19(2), 382–384 (1994).

66 Cantagrel A, Navaux F, Loubet-Lescoulie P *et al.* Interleukin-1β, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis. *Arthritis Rheum.* 42(6), 1093–1100 (1999).

67 Buchs N, Di Giovine FS, Silvestri T, Vannier E, Duff GW, Miossec P. IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: interaction with their plasma levels. *Genes Immun.* 2(4), 222–228 (2001).

68 Pawlik A, Kurzawski M, Florczak M, Gawronska Szklarz B, Herczynska M. IL1β+3953 exon 5 and IL-2 -330 promoter polymorphisms in patients with rheumatoid arthritis. *Clin. Exp. Rheumatol.* 23(2), 159–164 (2005).

69 Harrison P, Pointon JJ, Chapman K, Roddam A, Wordsworth BP. Interleukin-1 promoter region polymorphism role in rheumatoid arthritis: a meta-analysis of IL-1B-511A/G variant reveals association with rheumatoid arthritis. *Rheumatology (Oxford)* 47(12), 1768–1770 (2008).

70 Johnsen AK, Plenge RM, Butty V *et al.* A broad analysis of IL1 polymorphism and rheumatoid arthritis. *Arthritis Rheum.* 58(7), 1947–1957 (2008).

71 Srirangan S, Choy EH. The role of interleukin 6 in the pathophysiology of rheumatoid arthritis. *Ther. Adv. Musculoskelet. Dis.* 2(5), 247–256 (2010).

72 Huizinga TW, Keijsers V, Yanni G *et al.* Are differences in interleukin 10 production associated with joint damage? *Rheumatology (Oxford)* 39(11), 1180–1188 (2000).

73 Paradowska-Gorycka A, Trefler J, Maciejewska-Stelmach J, Lacki JK. Interleukin-10 gene promoter polymorphism in Polish rheumatoid arthritis patients. *Int. J. Immunogenet.* 37(4), 225–231 (2010).

74 Pawlik A, Kurzawski M, Szklarz BG, Herczynska M, Drozdzik M. Interleukin-10 promoter polymorphism in patients with

rheumatoid arthritis. *Clin. Rheumatol.* 24(5), 480–484 (2005).

75 Nemec P, Pavkova-Goldbergova M, Gatterova J, Fojtik Z, Vasku A, Soucek M. Association of the -1082 G/A promoter polymorphism of interleukin-10 gene with the autoantibodies production in patients with rheumatoid arthritis. *Clin. Rheumatol.* 28(8), 899–905 (2009).

76 Mcinnes IB, Liew FY. Cytokine networks – towards new therapies for rheumatoid arthritis. *Nat. Clin. Pract. Rheumatol.* 1(1), 31–39 (2005).

77 Baslund B, Tvede N, Danneskiold-Samsoe B *et al.* Targeting interleukin-15 in patients with rheumatoid arthritis: a proof-of-concept study. *Arthritis Rheum.* 52(9), 2686–2692 (2005).

78 Knevel R, Krabben A, Brouwer E *et al.* Genetic variants in IL15 associate with progression of joint destruction in rheumatoid arthritis: a multicohort study. *Ann. Rheum. Dis.* 71(10), 1651–1657 (2012).

79 Wajant H, Henkler F, Scheurich P. The TNF-receptor-associated factor family: scaffold molecules for cytokine receptors, kinases and their regulators. *Cell Signal.* 13(6), 389–400 (2001).

80 Wang Y, Rollins SA, Madri JA, Matis LA. Anti-C5 monoclonal antibody therapy prevents collagen-induced arthritis and ameliorates established disease. *Proc. Natl Acad. Sci. USA* 92(19), 8955–8959 (1995).

81 Plant D, Thomson W, Lunt M *et al.* The role of rheumatoid arthritis genetic susceptibility markers in the prediction of erosive disease in patients with early inflammatory polyarthritis: results from the Norfolk Arthritis Register. *Rheumatology (Oxford)* 50(1), 78–84 (2011).

82 Viatte S, Plant D, Lunt M *et al.* Investigation of rheumatoid arthritis genetic susceptibility markers in the early rheumatoid arthritis study further replicates the TRAF1 association with radiological damage. *J. Rheumatol.* 40(2), 144–156 (2012).

83 Kurreeman FAS, Padyukov L, Marques RB *et al.* A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med.* 4(9), e278 (2007).

84 Knevel R, De Rooy DP, Gregersen PK *et al.* Studying associations between variants in *TRAF1-C5* and *TNFAIP3-OLIG3* and the progression of joint destruction in rheumatoid arthritis in multiple cohorts. *Ann. Rheum. Dis.* 71(10), 1753–1755 (2012).

85 Kawabe T, Naka T, Yoshida K *et al.* The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity* 1(3), 167–178 (1994).

86 Van Der Linden MPM, Feitsma AL, Le Cessie S *et al.* Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis. *Arthritis Rheum.* 60(8), 2242–2247 (2009).

87 Naredo E, Collado P, Cruz A *et al.* Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. *Arthritis Rheum.* 57(1), 116–124 (2007).

88 Jain M, Samuels J. Musculoskeletal ultrasound as a diagnostic and prognostic tool in rheumatoid arthritis. *Bull. NYU Hos. Joint Dis.* 69(3), 215–219 (2011).

89 Taylor PC, Steuer A, Gruber J *et al.* Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. *Arthritis Rheum.* 50(4), 1107–1116 (2004).

90 Freeston JE, Wakefield RJ, Conaghan PG, Hensor EMA, Stewart SP, Emery P. A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power Doppler ultrasound when added to conventional assessment tools. *Ann. Rheum. Dis.* 69(2), 417–419 (2010).

91 Filer A, De Pablo P, Allen G *et al.* Utility of ultrasound joint counts in the prediction of rheumatoid arthritis in patients with very early synovitis. *Ann. Rheum. Dis.* 70(3), 500–507 (2011).

92 Hetland ML, Ejbjerg B, Horslev-Petersen K *et al.* MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). *Ann. Rheum. Dis.* 68(3), 384–390 (2009).

93 Mcqueen FM, Benton N, Perry D *et al.* Bone edema scored on magnetic resonance imaging scans of the dominant carpus at presentation predicts radiographic joint damage of the hands and feet six years later in patients with rheumatoid arthritis. *Arthritis Rheum.* 48(7), 1814–1827 (2003).

94 Palosaari K, Vuotila J, Takalo R *et al.* Bone oedema predicts erosive progression on wrist MRI in early RA – a 2-yr observational MRI and NC scintigraphy study. *Rheumatology (Oxford)* 45(12), 1542–1548 (2006).

95 Mohammed FF, Smookler DS, Khokha R. Metalloproteinases, inflammation, and rheumatoid arthritis. *Ann. Rheum. Dis.* 62(Suppl. 2), S43–S47 (2003).

96 Garnero P, Gineyts E, Christgau S, Finck B, Delmas PD. Association of baseline levels of urinary glucosyl–galactosyl–pyridinoline and type II collagen C-telopeptide with progression of joint destruction in patients with early rheumatoid arthritis. *Arthritis Rheum.* 46(1), 21–30 (2002).

97 Kobayashi Y, Udagawa N, Takahashi N. Action of RANKL and OPG for osteoclastogenesis. *Crit. Rev. Eukaryot. Gene Expr.* 19(1), 61–72 (2009).

98 Young-Min S, Cawston T, Marshall N *et al.* Biomarkers predict radiographic progression in early rheumatoid arthritis and perform well compared with traditional markers. *Arthritis Rheum.* 56(10), 3236–3247 (2007).

99 Posthumus MD, Limburg PC, Westra J *et al.* Serum levels of matrix metalloproteinase-3 in relation to the development of radiological damage in patients with early rheumatoid arthritis. *Rheumatology (Oxford)* 38(11), 1081–1087 (1999).

100 Yamanaka H, Matsuda Y, Tanaka M *et al.* Serum matrix metalloproteinase 3 as a predictor of the degree of joint destruction during the six months after measurement, in patients with early rheumatoid arthritis. *Arthritis Rheum.* 43(4), 852–858 (2000).

101 Garnero P, Landewe R, Boers M *et al.* Association of baseline levels of markers of bone and cartilage degradation with long-term progression of joint damage in patients with early rheumatoid arthritis: the COBRA study. *Arthritis Rheum.* 46(11), 2847–2856 (2002).

102 Hashimoto J, Garnero P, Van Der Heijde D *et al.* A combination of biochemical markers of cartilage and bone turnover, radiographic damage and body mass index to predict the progression of joint destruction in patients with rheumatoid arthritis treated with disease-modifying anti-rheumatic drugs. *Modern Rheumatol.* 19(3), 273–282 (2009).

103 Van Tuyl LHD, Voskuyl AE, Boers M *et al.* Baseline RANKL:OPG ratio and markers of bone and cartilage degradation predict annual radiological progression over 11 years in rheumatoid arthritis. *Ann. Rheum. Dis.* 69(9), 1623–1628 (2010).

104 Brennan P, Harrison B, Barrett E *et al.* A simple algorithm to predict the development of radiological erosions in patients with early rheumatoid arthritis: prospective cohort study. *BMJ* 313(7055), 471–476 (1996).

105 Drossaers-Bakker KW, Zwinderman AH, Vliet Vlieland TPM *et al.* Long-term outcome in rheumatoid arthritis: a simple algorithm of baseline parameters can predict radiographic damage, disability, and disease course at

12-year followup. *Arthritis Rheum.* 47(4), 383–390 (2002).

106 Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. *Rheumatology (Oxford)* 48(9), 1114–1121 (2009).

107 Visser K, Goekoop-Ruiterman YP, de Vries-Bouwstra JK *et al.* A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis receiving different dynamic treatment strategies: *post hoc* analyses from the BeSt study. *Ann. Rheum. Dis.* 69(7), 1333–1337 (2010).

▪ **Outlines a matrix risk model for rapid radiographic progression using routinely available clinical variables. It does, however, require evaluating in a separate cohort of rheumatoid arthritis patients having been developed and validated in a single randomized controlled trial.**

Best Practice & Research Clinical Rheumatology 25 (2011) 447–468

Contents lists available at SciVerse ScienceDirect

Best Practice & Research Clinical Rheumatology

journal homepage: www.elsevierhealth.com/berh

1

Precipitating and perpetuating factors of rheumatoid arthritis immunopathology – linking the triad of genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis

I.C. Scott, MBChB, MRCP, MSc [a,d,*], S. Steer, BA, MBBS, MSc, PhD, FRCP [c], C.M. Lewis, BA, MSc, PhD [d], A.P. Cope, BSc, PhD, FRCP, FHEA [a,b]

[a] *Department of Rheumatology, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK*
[b] *Department of Rheumatology, King's College School of Medicine, King's College London, Centre for Molecular and Cellular Biology of Inflammation, 1st Floor, New Hunt's House, Guy's Campus, Great Maze Pond, London SE1 1UL, UK*
[c] *Department of Rheumatology, King's College London, 3rd Floor, Weston Education Centre, Cutcombe Road, London SE5 9RJ, UK*
[d] *Department of Medical and Molecular Genetics, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK*

Keywords:
Rheumatoid arthritis
Environmental exposure
Genetics
Autoimmunity

Rheumatoid arthritis (RA) is considered to occur when genetic and environmental factors interact to trigger immunopathological changes and consequently an inflammatory arthritis. Over the last few decades, epidemiological and genetic studies have identified a large number of risk factors for RA development, the most prominent of which comprise cigarette smoking and the shared epitope alleles. These risks appear to differ substantially between anti-cyclic citrullinated peptide (ACPA)-positive and ACPA-negative disease. In this article, we will summarise the risk factors for RA development that have currently been identified, outlining the specific gene–environment and gene–gene interactions that may occur to precipitate and perpetuate autoimmunity and RA. We will also focus on how this knowledge of risk factors for RA may be implemented in the future to identify individuals at a high risk of disease development in whom preventative strategies may be undertaken.

* Corresponding author. Department of Medical and Molecular Genetics, King's College London, 8th Floor Tower Wing, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK. Tel.: +20 7188 7188; fax: +20 7407 7532.
E-mail address: ian.c.scott@btinternet.com (I.C. Scott).

1521-6942/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.berh.2011.10.010

Background

Rheumatoid arthritis (RA) is considered to occur when genetically predisposed individuals are exposed to specific environmental risk factors. These genetic and environmental risks interact to trigger perturbations in the immune system, with auto-antibody – rheumatoid factor (RF) and anti-cyclic citrullinated peptide antibody (ACPA) – generation in the majority of cases, followed by pro-inflammatory cytokine production and a consequent inflammatory arthritis.

Over the last few decades, epidemiological studies have identified a large number of environmental risk factors for RA. More recently, advances in genomics research have greatly increased our understanding of the genetic architecture underlying RA development, with over 30 risk alleles identified for sero-positive disease in individuals of European ancestry [1].

There is however a growing appreciation that RA, as opposed to being a single disease entity, is a syndrome comprising several distinct phenotypes [2,3]. The best-appreciated subdivision is by the presence or absence of immune responses to citrullinated protein antigens, termed ACPA-positive and ACPA-negative RA. Not only do these disease subtypes differ clinically, with ACPA-positive RA having higher rates of erosions and lower remission rates [4], but they also vary with regard to the genetic and environmental risk factors that contribute to their development [5].

In this article, we will summarise the genetic and environmental risk factors for RA that have been identified to date. We will outline how these factors may interact to precipitate and perpetuate autoimmunity and RA, focussing on the specific gene–environment interaction between smoking/periodontitis and the shared epitope alleles in the pathogenesis of ACPA-positive RA. Finally, we will describe how knowledge regarding RA risks may be implemented to identify and prevent RA development in high risk individuals in the future.

Genetic risk factors for RA development

Genetic factors dominate an individual's risk of developing RA, accounting for approximately two-thirds of the overall risk burden for both ACPA-positive and ACPA-negative disease [6]. RA is considered a complex polygenic disease, with multiple alleles contributing towards its development. Although the risk conferred by each individual risk allele is small, if several risk loci are present in the same individual they may be highly influential. To date, the majority of genome-wide associated studies (GWASs) have focussed on ACPA-positive disease; there is limited information regarding the genetic basis for ACPA-negative disease.

Genetic risk factors for ACPA-Positive RA

The majority of genetic risk for sero-positive RA is derived from the major histocompatibility complex, class II, DR beta 1 (HLA-DRB1). This is a group of alleles that encode the HLA class II DRβ-chain, which plays a pivotal role in antigen presentation by influencing the binding and presentation of arthritogenic peptides to auto-reactive $CD4^+$ T cells [7]. Although multiple HLA-DRB1 alleles – particularly DRB1*0401 and *0404 – are associated with RA, they all share a region of structural similarity termed the shared epitope (SE) [8]

At present over 30 non-MHC risk alleles for ACPA-positive RA have been identified and validated through candidate gene studies and GWAS [1]. The most prominent and well understood of these comprise variants of the *PTPN22* and *PADI4* genes. The *PTPN22* allele, 1858T, encodes the lymphoid-specific tyrosine phosphatase, Lyp, which is a negative regulator of T cell antigen receptor (TCR) signal transduction during T cell activation [9]. The variant associated with RA is a gain-of-function mutation that probably predisposes to autoimmune disease through excessive suppression of TCR signalling during thymic development, resulting in the survival of auto-reactive T cells [10]. *PADI4* is a peptidylarginine deiminase enzyme that post-transcriptionally converts arginine residues to citrulline [11]. It may therefore play a significant role in the development of ACPA through influencing protein citrullination.

Other important contributory loci are shown in Table 1, which summarises published odds ratios (ORs) of validated RA risk alleles from a recent meta-analysis of GWAS RA risk loci in sero-positive patients of European ancestry [1].

Genetic risk factors for ACPA-Negative RA

In comparison with ACPA-positive RA, there is a paucity of information on the genetic risks for ACPA-negative disease. Only a few studies exist that have examined the genetic basis of this disease subset. As with ACPA-positive disease, the HLA domain appears to be a potential area of importance, albeit one that confers lower risks in ACPA-negative RA. The non-shared epitope HLA-DRB1 alleles, DRB1*13 and DRB1*03, in combination, have been shown to be associated with ACPA-negative disease [12]. Additionally, the DRB1*13 allele appears to further contribute to ACPA-negative RA development by neutralising the effects of the SE alleles.

Variants in C-type lectin domain family 16, member A (*CLEC16A*), have also been shown in a candidate gene study to confer susceptibility to ACPA-negative but not ACPA-positive RA [13]. Although the function of this gene is unknown, it is almost exclusively expressed in immune cells and is associated with other autoimmune diseases such as multiple sclerosis, indicating an important biological role in autoimmunity.

Table 1
Validated risk alleles associated with sero-Positive RA in individuals of European ancestry [1].

	Locus	SNP ID	Candidate Gene(s)	OR (95% CI)
Established Risk Alleles	1p36	rs3890745	*TNFRSF14*	0.89 (0.85–0.94)
	1p13	rs2476601	*PTPN22*	1.94 (1.81–2.08)
	1p13	rs11586238	*CD2, CD58*	1.13 (1.07–1.19)
	1q23	rs12746613	*FCGR2A*	1.13 (1.06–1.21)
	1q31	rs10919563	*PTPRC*	0.88 (0.82–0.94)
	2p16	rs13031237	*REL*	1.13 (1.07–1.18)
	2q11	rs10865035	*AFF3*	1.12 (1.07–1.17)
	2q32	rs7574865	*STAT4*	1.16 (1.10–1.23)
	2q33	rs1980422	*CD28*	1.12 (1.06–1.18)
	2q33	rs3087243	*CTLA4*	0.87 (0.83–0.91)
	4q27	rs6822844	*IL2, IL21*	0.90 (0.84–0.95)
	6p21	rs6910071	*HLA-DRB1 (*0401 tag)*	2.88 (2.73–3.03)
	6q21	rs548234	*PRDM1*	1.10 (1.05–1.16)
	6q23	rs10499194	*TNFAIP3*	0.91 (0.87–0.96)
	6q23	rs6920220	*TNFAIP3*	1.22 (1.16–1.29)
	6q23	rs5029937	*TNFAIP3*	1.40 (1.24–1.58)
	6q25	rs394581	*TAGAP*	0.91(0.87–0.96)
	8p23	rs2736340	*BLK*	1.12 (1.07–1.18)
	9p13	rs2812378	*CCL21*	1.10 (1.05–1.16)
	9q33	rs3761847	*TRAF1, C5*	1.13 (1.08–1.18)
	10p15	rs2104286	*IL2RA*	0.92 (0.87–0.97)
	10p15	rs4750316	*PRKCQ*	0.87 (0.82–0.92)
	11p12	rs540386	*TRAF6*	0.88 (0.83–0.94)
	12q13	rs1678542	*KIF5A, PIP4K2C*	0.91 (0.87–0.96)
	20q13	rs4810485	*CD40*	0.85(0.80–0.90)
	22q12	rs3218253	*IL2RB*	1.09 (1.03–1.15)
Recently Validated Risk Alleles	2p14	rs934734	*SPRED2*	1.13 (1.06–1.21)
	5q11	rs6859219	*ANKRD55, IL6ST*	0.85 (0.78–0.93)
	5q21	rs26232	*C5orf30*	0.93 (0.88–0.98)
	3p14	rs13315591	*PXK*	1.13 (1.04–1.23)
	4p15	rs874040	*RBPJ*	1.18 (1.12–1.24)
	6q27	rs3093023	*CCR6*	1.11 (1.06–1.16)
	7q32	rs10488631	*IRF5*	1.25 (1.14–1.37)
	2q11	rs11676922	*AFF3*	1.15 (1.10–1.20)
	9p13	rs951005	*CCL21*	0.87 (0.81–0.93)
	10p15	rs706778	*IL2RA*	1.11 (1.06–1.17)

A further putative risk allele for ACPA-negative disease is interferon regulatory factor 5 (*IRF5*), a factor that induces interferon-alpha (IFN-α) transcription. Sigurdsson and colleagues examined its potential relationship to RA; they found that four single-nucleotide polymorphisms (SNPs) in the 5′ region of *IRF5* were associated with ACPA-negative disease in substantially sized cohorts of Swedish and Dutch RA cases and controls [14]. Although this indicates a likely role for *IRF5* and IFN-α in ACPA-negative RA pathogenesis, their findings were not reproduced in other European cohorts [15]. Further confirmatory research is therefore required.

To date, only a single GWAS has been performed that specifically examines genetic risks in ACPA-negative RA [16]. It evaluated genetic associations in 774 ACPA-negative patients and 1079 controls, which are relatively small numbers of participants when compared with GWAS performed in seropositive RA patients. Whilst no SNP achieved genome-wide association significance, the study had limited power and possible associations were seen for two new candidate loci, *RPS12P4* and *IGFBP1*, alongside the previously described *IRF5* locus.

Genetic risk factors for RA severity

In addition to being risk factors for disease susceptibility, there is also evidence that genes play important roles in determining disease phenotype and severity. Van der Helm-van Mil et al. examined this possibility by evaluating the variability in radiological hand damage in unrelated patients, monozygotic and dizygotic twins. They found that the variance in joint destruction was highest between unrelated patients, followed by dizygotic and finally monozygotic twins [17]. Although this study was too small to calculate heritability rates, it provides support for genetics as a determinant of RA severity. Further studies, outlined below, have examined the relationship between specific genetic loci (in particular, those that are RA susceptibility markers) and RA severity with some success.

HLA-DRB1 SE

The HLA-DRB1 SE alleles have been shown to associate with disease severity, with one retrospective literature review finding the HLA-DRB1*0401/*0404 genotype to be associated with a higher risk of early-onset RA with a more severe phenotype [18]. Similarly, Wagner et al. identified the same SE allele genotype to be a genetic marker for radiological damage in RA, being associated with higher rates of X-ray damage both at presentation and in more established disease [19]. Although these findings implicate the SE alleles to be predictors of disease severity, the impact of the SE may be secondary to its association with ACPA acting as a confounding variable.

PTPN22

There is some limited evidence that *PTN22* may contribute to disease severity. In a cross-sectional evaluation of 964 RA patients, Marinou et al. reported a trend towards higher rates of X-ray damage in RA patients carrying the *PTPN22* risk allele compared to those without it [20]. This association was however weak and other studies have failed to replicate their findings [21].

IL-1 locus

IL-1 is an important pro-inflammatory cytokine in RA, contributing to systemic inflammation and articular damage through the stimulation of local fibroblast proliferation and the activation of chondrocytes. Several studies have shown an association of polymorphisms at the IL-1 locus with radiological damage in RA. Cantagrel et al. found that a polymorphism in IL-1beta exon 5 enabled prognostic stratification for erosive disease with a high specificity in 42% of RA patients [22]. This association has been replicated in other studies [23].

TRAF1/C5 locus

The *TRAF1/C5* locus is an RA risk factor encoding complement component 5 and tumour necrosis factor (TNF)-receptor associated factor 1, which are important components of the inflammatory cascade. Polymorphisms at this locus have also been shown to be associated with the risk of developing radiological erosions in the Norfolk Arthritis Register (NOAR) cohort by Plant et al. [24]

CD40

CD40 encodes a protein that as a member of the TNF-receptor superfamily is important in a broad range of immune responses, e.g., memory B-cell development. It is a known genetic risk factor for RA development. More recently, the SNP, rs4810485, at the *CD40* locus has been shown to be associated with RA severity in the form of radiographic progression rates; this association has been validated and replicated in two cohorts [25].

Environmental risk factors for RA development

Many environmental factors have been shown in epidemiological case–control and cohort studies to be associated with the development of RA. The single risk factor that has an unequivocal association is smoking, which has been repeatedly shown in a variety of cohorts to increase the risk of sero-positive RA. Other robust associations comprise female gender, age, alcohol consumption and periodontitis. Many other environmental risk factors with weaker supporting evidence have also been linked to RA development; these comprise oral contraceptive pill use, vitamin D intake, previous blood transfusions, obesity, socio-economic status, non-inherited maternal antigens, breast-feeding and birth weight. Examples of these risk factors are shown in Table 2. As with genetic associations, environmental risk factors for RA appear to differ between ACPA-positive and ACPA-negative disease. Pederson highlighted these differences finding smoking and alcohol intake to be solely associated with ACPA-positive disease, and obesity to be predominantly linked to ACPA-negative disease [5].

Cigarette smoking

Cigarette smoking is strongly associated with RA. Its impact however appears limited to individuals with sero-positive disease [26,27]. Its influence is also greater in males compared with females [28]. This relationship is highlighted in a recent meta-analysis by Sugiyama et al. [29], who found that in 16 observational studies the summary odds ratio (OR) for developing RA was 1.89 (95% confidence interval (CI) 1.56–2.28) for males who had ever smoked and 1.27 (95% CI 1.12–1.44) for females who had ever smoked. The risks were higher in individuals possessing RF with sero-positive males who had ever smoked having an OR of 3.02 (95% CI 2.35–3.88) for developing RA.

Smoking has a dose-dependent effect on the risk of RA. Data from the Nurses' Health Study (NHS) found a linear relationship between RA and increasing pack-years of smoking with the age adjusted relative risk (RR) in females with an over 40 pack-year history of smoking comprising 1.99 (1.57–2.53), compared with 1.08 (0.86–1.36) for a 1–10 pack-year history [30]. This dose-dependent relationship is shown in Fig. 1.

Table 2
Some environmental risk factors for rheumatoid arthritis.

Risk factor	Odds ratio/Relative risk (95% confidence interval)
Ever smoked [29]	1.40 (1.25–1.58)
High alcohol intake [38]	0.50 (0.40–0.6)
Obesity [5]	1.57 (1.01–2.44)
High vitamin D intake [42]	0.67 (0.44–1.00)
High birthweight [49]	2.10 (1.40–3.30)

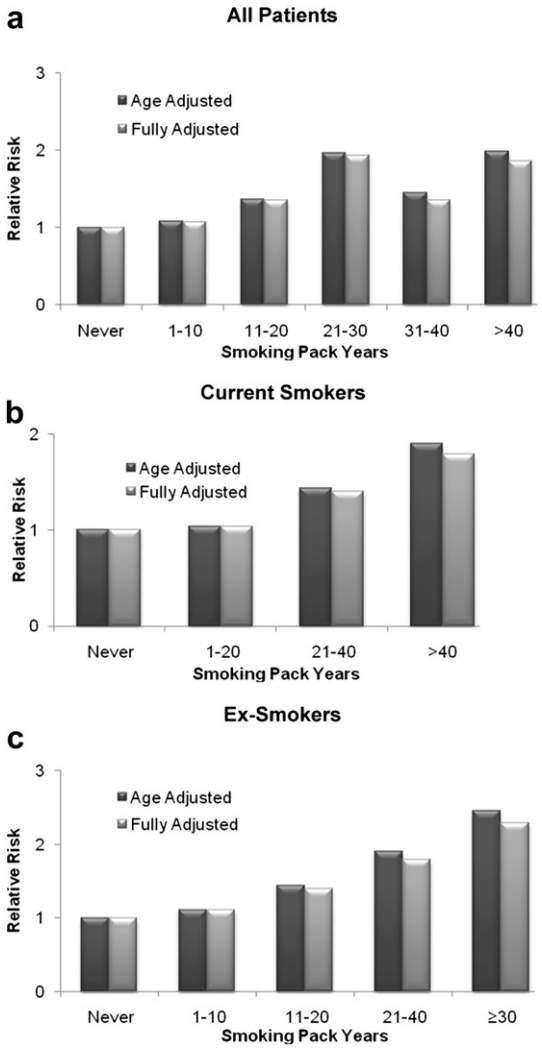


Fig. 1. Relative risk of rheumatoid Arthritis In Women from the Nurses' health study Stratified by Pack-years of smoking. a) all patients, b) current smokers, c) ex-Smokers. Figure adapted using data from Costenbader et al. [30].

Periodontitis

RA is prevalent in individuals with periodontitis [31]. A cross-sectional survey of 4461 US civilians who had undergone recent musculoskeletal and dental examinations found that individuals classified with RA were far more likely to be edentulous (OR 2.27; 95% CI 1.56–3.31) and have periodontitis (OR 1.82; 95% CI 1.04–3.20) compared with non-RA subjects. This association is greatest in those with seropositive RA [32].

Female gender

RA has a predilection for women, being on average 3 times more common in females compared with males. This relationship is highest in younger age groups, with the incidence of RA 5 times higher in Oslo females under the age of 50 years compared with age-matched males, but only twice as high in

those aged over 60 years [33]. The mechanism by which female gender increases RA susceptibility remains elusive. This gender association may represent sex hormone differences with supporting evidence arising from the fact that RA risk significantly increases in the post-partum period, with an OR of 5.6 (95% CI 1.8–17.6) for developing RA during the first 3 months post-partum [34].

Ageing

Although RA can affect any age group, its onset has reproducibly been shown to peak during the sixth decade of life [35,36]. This may result from immune senescence with the decline in host immunity with advancing age promoting immune reactivity to self-antigens [37]

Alcohol consumption

Alcohol intake appears protective against the onset of RA. Källberg and colleagues undertook a combined analysis of two independent case–control studies of RA, the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and the Case-Control study in Rheumatoid Arthritis (CACORA) [38]. They found that individuals consuming the highest levels of alcohol (defined as $\geq$5 drinks/week) had half the risk of developing RA in comparison with those consuming little or no alcohol. The risk reduction was greatest in individuals who were ACPA-positive and/or harbouring the SE. As with smoking exposure, the protective effects conferred by alcohol intake were dose-dependent with increasing levels of ethanol consumption further lowering the risk of RA. The dose-dependent effect of alcohol on RA risk is shown in Fig. 2.

Oral contraceptive pill use

The predilection of RA for females has led several research groups to examine the impact of oral contraceptive pill (OCP) use – and thus oestrogen exposure – on RA risk. The evidence to date is uncertain with some studies finding an increased risk of RA with OCP use, others finding OCP use to protect against RA development and some studies failing to identify any relationship between this risk factor and RA. A Swedish case–control study found that OCP use for $\geq$7 years reduced the risk of RA development (OR 0.37; 95% CI 0.15–0.93) [39]. By contrast, a Danish study found that OCP use increased the risk of ACPA-positive RA (OR 1.65; 95% CI 1.06–2.57) [5]. Pikwer et al. found no relationship between RA and OCP use in 136 women with RA and 544 age-matched controls [40]. Although in view of the strong gender association, it seems plausible that OCP use may influence the risk of RA, further work is required to better clarify this potential relationship.

Obesity

Obesity increases the risk of RA development. Its impact on RA risk however appears confined to ACPA-negative disease. Being morbidly obese (defined as a body mass index (BMI) of $\geq$30) a decade prior to RA onset has been shown to triple the risk of developing ACPA-negative RA (OR 3.45; 95% CI 1.73–6.87) [5]

High vitamin D intake

There is a growing appreciation of the role that vitamin D plays in autoimmunity, with its active form producing and maintaining immunological self-tolerance [41]. As a result, its relationship to RA development has been examined with contrasting outcomes; its precise impact on the risk of RA development remains uncertain.

Results from the Iowa women's health study, a prospective cohort study of 29 368 women aged 55–69 without RA at baseline, 152 of whom developed RA over 11 years of follow-up, found an inverse relationship between vitamin D intake and RA [42]. In this study, individuals with the highest level of vitamin D intake had a relative risk of 0.67 (95% CI 0.44–1.00) for RA when compared with

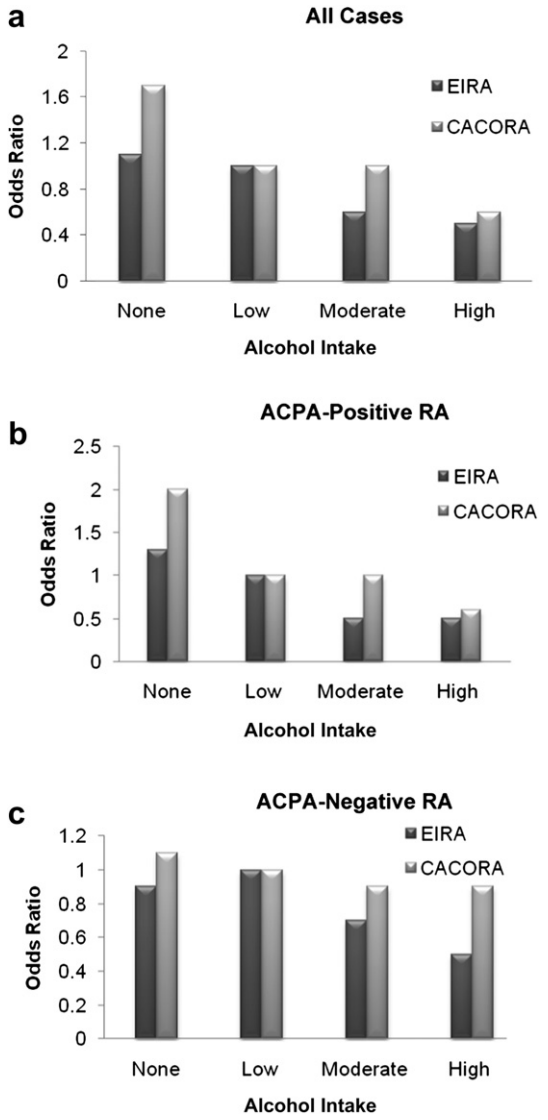


Fig. 2. Odds ratios for developing rheumatoid Arthritis In Individuals from the epidemiological investigation of rheumatoid arthritis (EIRA) and Case-Control study in rheumatoid arthritis (CACORA) studies Stratified by alcohol intake. a) all cases, b) ACPA-Positive RA, c) ACPA-Negative RA. Low alcohol intake = more than zero units but below or equal to the median level of consumption by controls; Moderate alcohol intake = more than median but below or equal to the 75th percentile of consumption by controls; high alcohol intake = above the 75th percentile of consumption of controls. All trends are significant apart from the ACPA-negative RA CACORA group. Figure adapted using data from Källberg et al. [38].

individuals with the lowest vitamin D intake. This relationship has not been reproduced in other cohorts [43].

Previous blood transfusion

Allogenic blood transfusions are known to have immunomodulatory effects. The exposure to foreign antigens in transfused blood has been shown to increase CD5 B cells alongside RF and

anti-nuclear antibody titres [44]. Examination of the impact of blood transfusions on RA risk in UK and US data sets has however provided opposing conclusions.

Analysis of cases and controls enrolled to the Norfolk Arthritis Register, a population-based study in Norfolk, England, found that previous blood transfusions increased the risk of RA (OR 3.58; 95% CI 1.46–8.81). The risks were approximately double in women (OR 4.36; 95% CI 1.45–13.06) compared with men (OR 2.69; 95% CI 0.43–16.75). The risks were also greater for sero-positive RA (OR 4.17; 95% CI 0.46–38.12) compared with sero-negative RA (OR 2.33; 95% CI 0.77–7.09) [45]. By contrast, examination of the US Iowa cohort gave opposing results with a history of blood transfusion being inversely associated with RA (RR 0.72; 95% CI 0.48–1.08) [46]. This difference may reflect variations in study methodology or international transfusion practices. Further work is required to clarify this relationship.

Socio-economic status

RA is more common in individuals from a lower socio-economic background. This finding has been highlighted in two independent case–control studies. Firstly EIRA, which found individuals without a university degree to have a RR of 1.4 (95% CI 1.2–1.8) for RA when compared to those with a university degree [47]. Similarly, in this cohort, the RR of sero-positive RA in non-manual employees compared with manual employees, assistant and intermediate non-manual employees grouped together was 1.5 (1.0–2.1), after adjustments for confounding variables such as smoking. These findings were reproduced in a Danish cohort, in which education levels were inversely associated with RA risk [48]. In this study, the RA risk in individuals with the longest formal education period was approximately half that of individuals with the lowest education levels.

Birth weight

High birth weight has been shown to be associated with RA. Analysis of the NHS cohort identified a birth weight of more than 4.5 kg to confer a twofold increased risk of RA compared with a lower birth weight of 3.2–3.85 kg (RR 2.1; 95% CI 1.4–3.3) [49]. The authors of this article considered that this relationship may arise from dysregulation of the hypothalamus–pituitary–adrenal (HPA) axis, with plasma cortisol levels being both inversely related to birth weight and abnormally low in adults with RA.

Breast-feeding

Breast-feeding has been shown in two large case–control studies to protect against RA. Karlson et al. examined the impact of breast-feeding on RA risk within the NHS cohort [50], finding that breast-feeding for more than a year was inversely related to RA development. As with other environmental risk factors, this protective effect appeared exposure dependent with a longer duration of breast-feeding providing an increasing trend towards risk reduction. For women who had breast-fed compared with parous women who had not breast-fed, the RR for RA comprised 0.5 (95% CI 0.3–0.8) when breast-feeding for $\geq$24 months; 0.8 (95% CI 0.6–1.0) for 12–23 months; 0.9 (95% CI 0.7–1.1) for 4–11 months; and 1.0 (95% CI 0.8–1.2) for $\leq$3 months. Similar results were found in a nested case–control study using information from the Malmö Diet and Cancer Study in Sweden [40]. In this cohort, the OR for RA in women who had breast-fed for $\geq$13 months compared with those who had never breast-fed was 0.46 (95% CI 0.24–0.91); in those who had breast-fed for 1–12 months the OR was 0.74 (95% CI 0.45–1.20).

The impact of breast-feeding on RA risk is however not completely defined. Brennan et al. reported opposing results in their case–control study finding breast-feeding to be more common in RA patients compared with age- and sex-matched controls [51]. This study was however of a smaller size and may have been confounded by HLA-DR4 status [52]. A further study failed to find any association between breast-feeding and the risk of RA development, although it did identify breast-feeding as a risk factor for more severe disease [53].

Non-inherited maternal antigens

Non-inherited maternal antigens (NIMAs) are antigens within the offspring that are encoded by maternal as opposed to inherited alleles (termed 'microchimerism'). They can be passed from the mother to her offspring from 3 months gestation until full term [54]. At the same position in the MHC molecule as the SE, the amino acid sequence 'DERRA' (comprising aspartic acid, glutamic acid, arginine, alanine and alanine) has been shown to protect against RA. This sequence is encoded by non-SE HLA-DRB1 alleles, that is, HLA-DRB1*0103. The OR of RA in people carrying HLA-DRB1 alleles expressing the 'DERRA' sequence compared with those without the SE or 'DERRA' alleles is 0.5–0.7 [55]. DERRA containing MHC molecules can be presented as NIMA. In these circumstances, they provide the same level of protection against RA as DERRA expressing alleles inherited directly from a parent.

RA immunopathology – an overview

RA is characterised by chronic systemic and articular inflammation. Although the precise immunopathological mechanisms that underlie RA have not been completely defined, this process is driven by both the innate and adaptive immune systems with T cells, B cells, macrophages, neutrophils and synovial fibroblasts playing important roles.

Macrophages

Macrophages and their precursor's monocytes act both systemically, through the production of classical RA pro-inflammatory cytokines, that is, IL-1 and TNF-α [56] and, locally, through synovial infiltration with macrophages/monocytes being enriched in the rheumatoid synovium and destructive pannus tissue [57].

T cells

T cells are well established as key components of RA immunopathology. Evidence for this stems from the strong association of RA with the SE alleles encoding the MHC (indicating that antigen presentation to T cells is an important process in RA), the prevalence of $CD4^+$ T cells within rheumatoid synovium and the efficacy of the selective T-cell co-stimulation modulator, abatacept [58]. Antigen-dependent T-cell responses may be important in initiating the inflammatory response during RA [59]. They may also act independently of antigenic stimulation to perpetuate inflammation through activating monocytes/macrophages to produce pro-inflammatory cytokines [60]. Additionally, T helper 17 cells produce IL17, which has pleiotropic effects on many RA effector cells causing inflammation and driving osteoclastogenesis and bone resorption [61].

B cells

The importance of B cells in RA is highlighted by the efficacy of B-cell depletion therapy with the anti-CD20 monoclonal antibody, rituximab [62]. B cells have multiple functions in RA. They can act as antigen-presenting cells (APCs), presenting antigens via MHC class II molecules to T cells activating them with consequent downstream macrophage activation and TNF-α production. They produce pro-inflammatory cytokines and chemokines directly via Toll-like receptor activation [63]. B cells are responsible for auto-antibody production, with RF and ACPA forming immune complexes with immune responses via Fc and complement receptors [64]. Additionally, in many RA patients, synovial extra-follicular germinal centres develop with B lymphocytes surrounded by T cells, acting as functional ectopic germinal centres [65].

Synovial fibroblasts

Fibroblast-like synoviocytes are prevalent in RA synovium where they have a unique phenotype with aggressive and invasive properties; they drive cartilage erosion through matrix metalloproteinase production and are dominant producers of IL-6 [66].

Gene–Environment interactions precipitating RA

The underlying paradigm of RA pathogenesis is that genetic and environmental risk factors interact to precipitate immune system changes, autoimmunity and, subsequently, disease. This concept clearly represents an oversimplification of the underlying disease aetiology with some of the identified RA risk factors increasing RA risk, some conferring protection against RA, and some interacting with one another to modify their impact on the development of RA.

Attempts to provide a unifying model in which all of these risks interact to precipitate RA have been to a large extent unsuccessful. However, with the advent of widespread ACPA serology availability, a clear gene–environment model has been revealed that delineates how two environmental risks (smoking and chronic periodontitis) may interact with the SE alleles to precipitate ACPA formation and the development of ACPA-positive RA. In this section, we will outline this biological model in detail, describing the evidence underlying it.

What are citrullinated peptides

Citrulline is a non-standard amino acid that results from post-translational modification of arginine resides [67]. This modification, termed 'citrullination' or 'deimination', is facilitated by a family of peptidylarginine deiminase (PAD) enzymes in a calcium-dependent manner. The substitution of arginine for citrulline results in key changes in the structure and ionic charge of the peptide from positive to neutral, with a consequent potential for functional differences.

Although citrullinated peptides are present in healthy individuals where they play important physiological roles aiding in epidermal cornification and myelin sheath insulation [68], they are far more abundant in individuals with inflammatory disorders including multiple sclerosis, myositis and non-RA inflammatory arthritidies [69,70].

Auto-immunity to citrullinated peptides in RA – evidence for a pathological role

Although citrullinated proteins are not confined to individuals with RA, the presence of antibodies that target them are almost exclusively present in this disease state. A meta-analysis of 86 studies showed ACPA had a sensitivity and specificity of 67% and 95% [71]. Although its sensitivity approximates that of immunoglobulin M (IgM) RF, it is substantially more specific.

The currently available ACPA assays are of limited use in attempting to understand disease pathogenesis because they incorporate a number of citrulline-containing peptides that fail to correspond with *in vivo* citrullinated proteins in the joint. There is, however, an increasing number of citrullinated peptides that have been established as antigens expressed within the articular environment. These include fibrinogen/fibrin, vimentin and α-enolase. There is ongoing work to identify further pathogenic citrullinated peptides.

The pathological role that ACPA appears to play in causing ACPA-positive RA is indicated by its presence in the serum of individuals with RA up to 14 years prior to the onset of clinical features of this disease. Nielen et al. analysed archived blood samples from 79 RA patients who had previously donated blood. They found that 49% of patients had positive ACPA serology 4.5 years pre-RA onset compared with less than 1% of 2138 controls [72]. Similarly, through evaluating a biobank of approximately 90 000 blood donors from Sweden that contained 83 incident cases of new-onset RA, Rantapää-Dahlqvist et al. found that ACPAs were significant predictors of RA occurring in 34% of pre-RA patients compared with only 2% of matched controls [73].

Environmental drivers of protein citrullination in individuals pre-RA

There is substantial evidence that two processes – smoking and periodontitis – are at least in part responsible for increased protein citrullination pre-RA.

Smoking, shown to be almost exclusively associated with ACPA-positive RA, appears to increase protein citrullination through promoting expression of PAD enzymes in alveolar cells. Evidence for this stems from finding substantial up-regulation of citrullinated proteins with enhanced PAD2 enzyme expression in bronchoalveolar lavage cells from smokers compared with non-smokers [74].

Furthermore, autoimmunity to citrulline may be promoted through increased thiocyanate ions produced by tobacco smoke metabolism [75]. Thiocyanate is metabolised to homocitrulline, which is similar in shape and structure to citrulline. Its increased presence in smokers may promote autoantibody formation with cross-reactivity to citrulline.

Periodontitis is a destructive inflammatory disease of the supporting tissues of the teeth. It is caused by specific microorganisms, one of the best characterised of which is *Porphyromonas gingivalis.* As previously discussed, periodontitis is prevalent in RA, particularly ACPA-positive disease. Of crucial importance is the fact that *P. Gingivalis* is the only known bacterium that expresses PAD. It may thus contribute to citrullinated peptide neoepitope generation in pre-RA individuals with periodontitis through PAD-induced protein citrullination. Further support for the pathogenic role of this microbe in anti-citrulline immunity comes from the finding that increased titres of anti-*P. Gingivalis* antibodies correlate with ACPA subtypes in RA patients [76].

Genetic influences on ACPA formation in individuals pre-RA

The crucial issue of how immune tolerance to citrulline could be breached in pre-RA individuals has been addressed by Hill and colleagues. They examined the T-cell response to citrullinated peptides in HLA-DRB1*0401 transgenic mice [7]. MHC class II molecules harbouring the SE contain a specific amino acid sequence within pocket 4 (P4) – a positively charged area influencing binding of antigenic peptides that favours negatively charged amino acid residues. Upon citrullination, the positively charged arginine is converted to citrulline, which lacks an ionic charge and has a better structural conformation with P4. Hill et al. identified that as a result of these structural and ionic alterations citrullinated peptides bound to the MHC P4 in mice possessing the SE with a 100-fold greater affinity when compared with their arginine-containing precursors. This peptide/SE complex was subsequently presented to $CD4^+$ T cells activating them. The ability of citrullinated peptides to induce a T-cell response has been replicated in human subjects by Feitsma et al., who found that naturally occurring citrullinated vimentin peptides were recognised by T cells from ACPA-positive HLA-DR+ RA patients [77].

The ability of ACPA to drive actual joint inflammation was subsequently shown in a mouse model [78]. HLA-DRB1*0401 transgenic mice were immunised with citrullinated human fibrinogen; their outcomes were compared to wild-type mice. Only mice possessing the HLA transgene developed an inflammatory arthritis, which directly implicates citrullinated fibrinogen as an arthritogenic peptide in the context of SE containing MHC class II molecules.

In summary, it appears probable that ACPA formation is triggered by citrullinated peptides binding with a high-affinity to SE containing MHC molecules. This peptide/SE complex is then presented to $CD4^+$ T cells, which activate B cells driving ACPA formation.

Progression from asymptomatic ACPA-Positive individuals to clinical RA

Because ACPA can predate clinical RA by many years alongside the fact that not all individuals with ACPA develop RA, it seems apparent that further factors are required to trigger the shift from being an asymptomatic individual with ACPA to having established RA with synovitis.

An important feature of pathogenic antibodies is that they possess fine specificity for certain antigens. It has been postulated that for ACPA to become pathogenic and elicit articular damage they need to fully mature and increase their number of antigen specificities. This would explain the latent period that exists between ACPA formation and RA onset. Evidence for this maturation process in RA, termed 'epitope-spreading', exists with Van der Woude et al. finding that the number of citrullinated antigens recognised by ACPA increased in the time period leading up to RA onset in individuals with an undifferentiated arthritis [79].

Although the precise mechanisms that stimulate epitope-spreading and the onset of RA remain elusive, it has been proposed that a second event, that is, infection or trauma occurs, which triggers a non-specific synovitis with associated citrullination within the joint [80]. In healthy individuals, this would resolve without sequela; however, in those possessing ACPA and T cells reactive to them this would lead to a chronic inflammatory arthritis evolving into RA. ACPA-mediated immune complexes

could subsequently drive macrophage TNF-α production alongside other pro-inflammatory cytokine pathways [81].

Evidence for the smoking-shared epitope interaction

Several prominent case–control studies have confirmed the significant gene–environment interaction between the SE alleles and smoking in ACPA-positive RA. This supports the concept that smoking-induced citrullinated peptides interact with the SE to induce autoimmunity and RA. These studies also show a cumulative genetic effect on RA risk with disease risks being greater in those possessing more copies of the SE alleles. This implies that a greater phenotypic expression of the SE increases the degree of citrullinated peptide presentation to $CD4^+$ T cells, driving RA development.

One key study to examine this important interaction evaluated a Swedish population-based case–control cohort [26]. The impact of different SE and smoking combinations on RA risk was compared to the RA risk in individuals who had never smoked and possessed no SE alleles. In this analysis, the RR of sero-positive RA was 2.4 (95% CI 1.3–4.6) in smokers with no SE genes, 5.5 (95% CI 3.0–10.0) in current smokers with one SE gene and 15.7 (95% CI 7.2–34.2) in current smokers with two copies of the SE genes. This multiplicative interaction is highlighted in Fig. 3.

This finding was reproduced within a Danish case–control study of 309 ACPA-positive and 136 ACPA-negative recent-onset RA cases alongside 533 sex-/age-matched controls [82]. In this cohort, the risks of ACPA-positive RA also increased exponentially in smokers depending on the number of SE gene copies they carried. In individuals with a >20 pack-year history of smoking the OR for RA was 1.22 (0.48–3.08) in SE non-carriers, 9.66 (4.38–21.3) in SE heterozygotes and 52.6 (18.0–154) in SE homozygotes.

Evidence for a gene–gene interaction between HLA-DR4 and PTPN22

The interaction between the SE and smoking is complicated by an additional outside influence from *PTPN22*. This gene interacts with the SE allele, modifying the risk of RA in smokers. Kallberg and colleagues examined three large case–control studies: EIRA, the North American RA Consortium (NARAC) study and the Leiden early arthritis cohort [83]. They found a significant interaction between

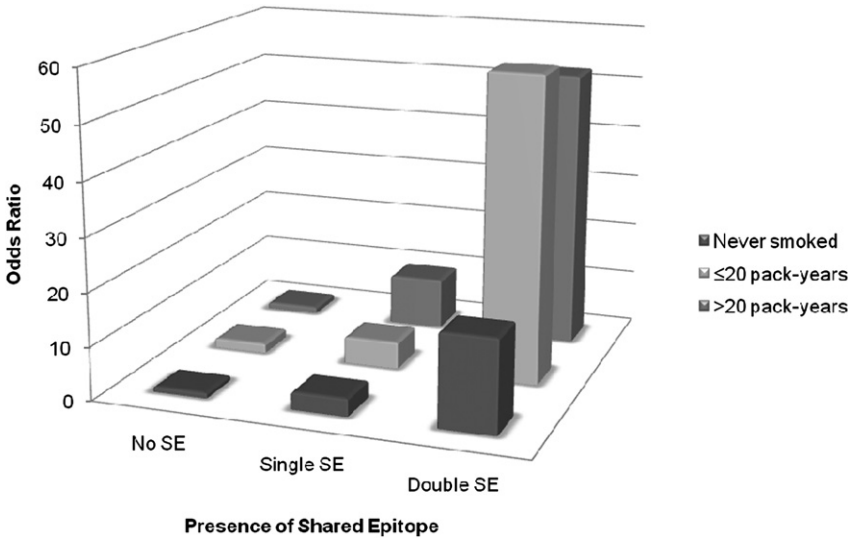


Fig. 3. The interaction between smoking and the number of shared epitope allele copies on the risk of ACPA-Positive rheumatoid arthritis. Figure adapted using data from Pedersen et al. [82].

HLA-DRB1 SE alleles and the *PTPN22* R620W allele in smokers. Although in the absence of the SE alleles *PTPN22* had no impact on RA risk, in the presence of the SE it substantially elevated the risk of ACPA-positive RA. The OR for RA in individuals who had ever smoked and who carried no *PTPN22*/a single SE allele was 5.4 (95% CI 3.3–8.9), with any *PTPN22*/a single SE allele was 11.3 (95% CI 6.2–20.5), with no *PTPN22*/a double SE allele was 23.6 (95% CI 12.8–43.8) and with any *PTPN22*/a double SE allele was 23.4 (95% CI 10.4–52.4). This interaction is shown in Fig. 4.

Hypothetical model for Gene–Environment interactions precipitating ACPA-positive RA

In summary, the following stages represent a biologically plausible model through which genetic and environmental factors interact to cause ACPA-positive RA (outlined in Fig. 5).

1. Peripheral protein citrullination occurs – either through periodontal infection with *P. gingivalis* or through smoking.
2. Structural and ionic changes resulting from protein citrullination enhance the binding of citrullinated self-antigens to MHC class II molecules in individuals possessing the SE alleles.
3. SE/citrullinated peptide complexes are presented by APCs to $CD4^+$ T cells, activating them.
4. These in turn activate B cells driving ACPA formation.
5. A second factor – for example, trauma or infection – precipitates synovial inflammation and the development of citrullinated proteins within the articular environment. These are targeted by ACPA.
6. ACPA–citrulline immune complexes are formed. These activate macrophages/monocytes driving inflammatory cytokine production and precipitating RA.

Combining genetic and environmental factors to identify individuals at a high risk of RA

An increased knowledge regarding the environmental and genetic risk factors that underlie RA development not only improves our understanding of RA pathogenesis but may, through the use of

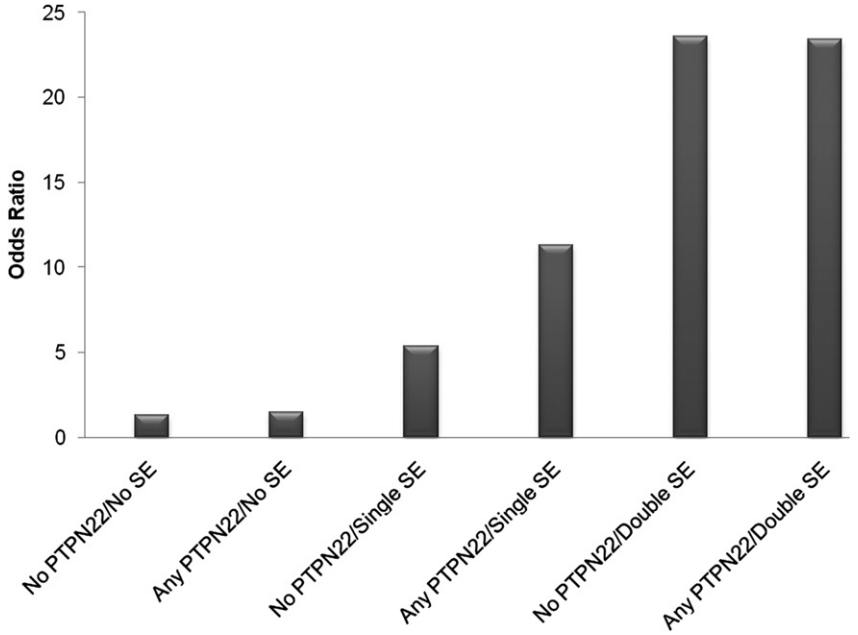


Fig. 4. Odds ratios for developing ACPA-Positive RA In Smokers provided by different HLA-DRB1 SE and Minor R620W *PTPN22* allele combinations. Figure adapted using data from Kallberg et al. [83].

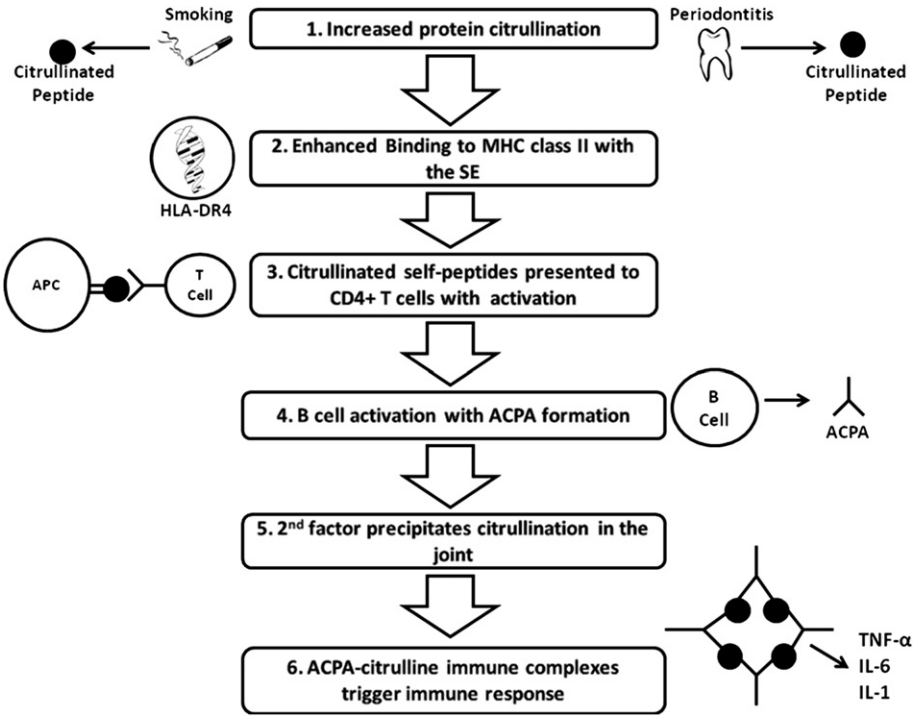


Fig. 5. Hypothetical model for gene–environment interactions precipitating ACPA-Positive RA.

prediction modelling, also facilitate the identification of people at a high risk of developing RA in the future. Such a prediction model would allow disease-prevention strategies to be evaluated in high risk individuals. Although each individual RA risk factor confers only a small risk, when used multiplicatively they may be highly informative. This is particularly relevant in *a priori* high-risk groups (e.g., individuals with ACPA antibodies or strong family histories of RA).

To date, only one RA prediction model has been successfully developed for use in the general asymptomatic population [84]. This model combined 22 RA risk alleles with the epidemiological risk factors smoking, ageing and female gender. It provided significant risk stratification: patients with the highest number of genetic risk factors had a three- to sixfold increased odds of sero-positive RA when compared to patients with an 'average' background genetic risk. The addition of genetic risk factors to a model that used clinical risks alone increased its accuracy improving the area under the curve, which is a commonly used measurement in risk-prediction models, from 0.63 to 0.75. Although this model provided some individuals with relatively high ORs of RA, owing to the relatively low lifetime risk of RA, their absolute risk of RA development remained small and the clinical utility of this prediction model was therefore limited. Its use would, however, be improved by incorporating factors providing greater risks of RA, which may become available with newer genotyping techniques. At present, further work is required to develop an accurate prediction model that influences clinical practice.

Preventing RA in high-risk individuals

Current RA treatment strategies focus on the titration of disease-modifying anti-rheumatic drugs (DMARDs) and biologic agents according to disease activity. This approach is limited, achieving remission in only 16–42% of cases [85]. There is, therefore, a key need to adopt different treatment approaches, one facet of which could be instituting disease-prevention treatments in individuals at a high risk of RA. Several small randomised controlled trials (RCTs) have evaluated this approach with some success.

Primary or secondary RA prevention

Bykerk classified RA-prevention techniques as belonging to two groups: (1) primary prevention – undertaken in individuals with genetic risks for RA in whom the pathogenic process has not yet started – and (2) secondary prevention – in individuals with pre-clinical disease either at an early asymptomatic stage or in the late pre-clinical stages when symptoms are present [86].

Because no robust system has been devised that can accurately identify asymptomatic individuals at a high risk of RA from the general population, it is currently not possible to undertake primary prevention strategies. As our knowledge of RA risks advances, this may however become an area of therapeutic potential.

Secondary prevention strategies have been instituted with some success. In these instances, patients with pre-clinical RA were identified by the presence of biological markers such as ACPA alongside early clinical manifestations associated with progression to RA. All studies examined the ability of pharmacological treatments comprising corticosteroids, methotrexate or biologics to prevent the progression of pre-clinical RA to clinically overt disease. To date, no studies have examined the impact of limiting exposure to environmental risks such as smoking in RA prevention.

Corticosteroids as secondary prevention

Several studies have shown some evidence that corticosteroids may play a role in RA prevention. Bos et al. evaluated 83 patients with arthralgia and ACPA/RF randomised to receive either intramuscular (IM) dexamethasone or placebo. Corticosteroid treatment reduced ACPA levels within 1 month (persisting for 6 months) and reduced disease activity scores at RA onset. Although this suggested that short-term corticosteroids attenuated the disease process, it had no impact on the number of patients progressing to synovitis ($\sim$20% in both treatment and placebo groups) [87]. Similarly, Verstappen and colleagues evaluated whether treating very early undifferentiated inflammatory polyarthritis (VEIA) with corticosteroids reduced the future requirement for DMARDs and RA development. Patients with VEIA were randomised to once weekly IM 80 mg depomederone injections for 3 weeks or placebo. At 12 months, the number of placebo-treated patients requiring DMARDs was twice that of those treated with corticosteroids [88]. These studies provide some support for a role of corticosteroids in RA prevention, the efficacy of which may be improved with more intensive and prolonged regimes. Further research is, however, required to better establish their role in RA secondary prevention pathways.

Methotrexate as secondary prevention

The PRObable rheumatoid arthritis: Methotrexate versus Placebo Treatment (PROMPT) study evaluated methotrexate as a preventative treatment for RA [89]. In this double-blind, placebo-controlled randomised controlled trial (RCT), 110 patients with an undifferentiated arthritis fulfilling the American College of Rheumatology (ACR) criteria for probable RA were treated with methotrexate or placebo. The starting dose of 15 mg/week methotrexate was titrated every 3 months until the DAS28 was $\leq$2.4; at 12 months, treatment was tapered and stopped. By the end of the study period (30 months), 40% of patients in the methotrexate group had progressed to RA compared with 53% of the placebo group ($P = 0.046$). Individuals treated with methotrexate also developed RA at a later time point and had less radiological progression than those treated with placebo. This trial provides some support for the use of methotrexate for RA prevention.

Biologics as secondary prevention

The ADJUST trial (Abatacept study to Determine the effectiveness in preventing the development of rheumatoid arthritis in patients with Undifferentiated inflammatory arthritis and to evaluate Safety and Tolerability) examined the impact of T-cell co-stimulation modulation on the development of RA in patients with undifferentiated arthritis or very early RA [90]. Patients were randomised to 6 months treatment with abatacept or placebo. At 2 years, a non-statistically significant reduction in individuals meeting the RA classification criteria was observed in those treated with abatacept versus placebo: 46%

of abatacept-treated patients were classified with RA compared with 67% of the placebo group. Placebo-treated patients also more frequently developed magnetic resonance imaging (MRI) structural changes.

The impact of infliximab on preventing RA development has also been evaluated in a small randomised study [91]. Seventeen patients with an undifferentiated arthritis of less than 12 months duration were randomised to infliximab or placebo for 14 weeks. Infliximab had no impact on progression to RA with 100% of patients in the treatment arm developing RA.

Further work – A research agenda

Although the last few decades have seen marked progress in our understanding of the risk factors for RA alongside it's pathogenesis and immunopathology there remain a large number of unanswered questions and key areas that require further evaluation. In this section, we will outline some of these areas, providing a focus for further research.

Clarify key environmental risk factors for RA

Although many environmental risk factors for RA have been identified, their associations are often weak, do not have a clear biological relationship to RA and – with the exception of smoking, ageing and gender – are often not reproducible. There is, therefore, a need to better clarify which environmental risk factors for RA are truly important. Due to the relative rarity of RA, cohort studies (the gold standard to examine disease risk factors) are difficult and time consuming to perform. There is therefore a large reliance on case–control studies to examine environmental risks for RA, which are open to impartiality in the form of recall and recruitment bias. It is vital to ensure that future case–control studies examining RA risk factors both recruit newly diagnosed RA patients to minimise recall bias and recruit controls from the same defined study population as cases. By optimising study design, the likelihood of finding true environmental risks for RA will be increased.

Identify the 'missing heritability' of RA

There is a need to better characterise the genetic basis of RA. Although over 30 genetic risk loci for RA have been found, a large amount of the genetic risk for RA remains unidentified with known genetic factors only explaining at most 40% of sero-positive RA heritability (37% from HLA; 5% from other loci) [92]. The remaining genetic risks may be identified through modern genotyping platforms that better tag underlying RA causal genetic variants with higher disease relative risks or large-scale sequencing that detect structural variations such as copy number variants and translocations, which may significantly contribute to RA development.

Examine genetic risks for RA across ethnic groups

There is a key need to establish the genetic basis of RA in different ethnic groups, with current GWAS in RA being limited to individuals of European ancestry [1]. The limited published data in this area indicate that important differences in genetic risks between ethnic groups exist, with Lee et al. reporting that none of the susceptibility loci in Caucasian RA patients contributed to disease in Koreans [93]. Although Hughes et al. found overlapping ORs and 95% CIs in 24 of 27 candidate SNPs between 556 sero-positive African Americans with RA and those of European ancestry, only one European risk allele had a statistically significant association in African Americans with RA [94].

Increase understanding of risk factors for ACPA-Negative RA

It is becoming increasingly apparent that ACPA-positive RA differs from ACPA-negative RA with regard to its genetic and environmental basis. To date, only one GWAS has been undertaken in ACPA-negative RA [16]. Additionally, many epidemiological studies have examined either sero-positive RA or RA subsets grouped together. It is therefore vital that future epidemiological and genetic studies

account for disease heterogeneity when examining risk factors for RA subsets, with increased attention given to evaluating ACPA-negative RA.

Increase knowledge of gene–gene and gene–environment interactions

There is a limited amount of information on the gene–environment and gene–gene interactions that exist between RA risk factors. Previous studies examining this have found key interactions between smoking/the SE alleles and the SE alleles/*PTPN22*. It is thus likely that other risk factor interactions exist, which can substantially modify their risks conferred for disease. Further research is required in this important area, which is of particular relevance in the development of RA prediction models to ensure their accuracy.

Indentify Genetic markers for disease severity

There is a need to better establish genetic markers for disease severity, with existing research showing heritability for disease outcomes such as radiographic erosions [17]. Current clinical parameters for identifying individuals at risk of developing severe RA are limited. A recent matrix risk model for rapid radiological progression using clinical factors had limited accuracy, with positive and negative predictive values of 62% and 91%, respectively [95]. This highlights a need for improved disease severity prediction markers to aid early prognostic stratification and allow prompt appropriate treatment to be tailored on an individual basis; this may be provided by genetic markers.

Develop and refine RA risk-prediction models

It is of crucial importance to use the accrued information regarding RA risks in clinical practice through developing and refining RA risk-prediction models. A clinically useful prediction model would allow further evaluation of preventative strategies, with short courses of corticosteroids, methotrexate and abatacept already showing some promise in preventing RA development in preclinical RA. It would also enable the investigation of high risk individuals to gain an increased understanding of the precise immunopathological changes that occur to precipitate this common and important disease.

Conclusions

RA is a heterogeneous disease, with ACPA-positive and ACPA-negative disease differing not only phenotypically but also with regard to the risk factors that underlie their development. Many environmental and genetic risks for RA have been identified; however, these are mainly limited to seropositive disease and individuals of European ancestry. Additionally, with the exception of smoking, age and female gender, many environmental risk factors have weak supporting evidence. Furthermore, the precise risk genes that underlie RA development remain uncertain, with approximately 60% of the genetic basis for RA unknown. These points highlight key needs for well-designed epidemiological studies to better clarify which environmental associations are important, improved genotyping techniques to identify the 'missing heritability' of RA, GWAS to be undertaken within different ethnic groups and, finally, a requirement for the risks for ACPA-positive and ACPA-negative RA to be evaluated separately. There is strong evidence for a gene–environment interaction between the SE alleles and smoking that precipitates ACPA generation, leading to ACPA-positive RA. By gaining an increased understanding of the risk factors for RA, we will increase our knowledge of disease subtype pathogenesis. This knowledge may also enable the development of RA prediction models that accurately identify those individuals at a high risk of RA, enabling disease-prevention strategies to be employed.

Funding and conflicts of interest

Dr Ian Scott and Professor Andrew Cope receive funding for their research into rheumatoid arthritis from Arthritis Research UK. None of the authors have a relevant conflict of interest.

Research agenda

- Better clarification of the environmental risk factors for RA
- Identification of the 'missing heritability' of RA
- Gaining an improved understanding of RA risk factors across ethnic groups and disease subtypes
- Improving the understanding of the genetic predictors for disease severity
- Further development and refinement of RA prediction models

References

*[1] Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nature Genetics 2010;42:508–16.
[2] van der Helm-van Mil AH, Huizinga TW. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. Arthritis Research and Therapy 2008;10:205.
[3] van Oosterhout M, Bajema I, Levarht EW, Toes RE, Huizinga TW, van Laar JM. Differences in synovial tissue infiltrates between anti-cyclic citrullinated peptide-positive rheumatoid arthritis and anti-cyclic citrullinated peptide-negative rheumatoid arthritis. Arthritis Rheumatology 2008;58:53–60.
[4] van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Toes RE, Huizinga TW. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. Arthritis Research and Therapy 2005;7:R949–58.
*[5] Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. Arthritis Research and Therapy 2006;8:R133.
[6] van der Woude D, Houwing-Duistermaat JJ, Toes RE, Huizinga TW, Thomson W, Worthington J, et al. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein-negative rheumatoid arthritis. Arthritis Rheumatology 2009;60:916–23.
*[7] Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, Cairns E. Cutting Edge: the Conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. Journal of Immunology 2003;171:538–41.
*[8] Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheumatology 1987;30:1205–13.
[9] Michou L, Lasbleiz S, Rat AC, Migliorinin P, Balsa A, Westhovens René, et al. Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. Proceedings of the National Academy of Sciences 2007; 104:1649–54.
[10] Bottini N, Vang T, Cucca F, Mustelin T. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. Seminars in Immunology 2006;18:207–13.
[11] Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nature Genetics 2003;34:395–402.
[12] Lundström E, Källberg H, Smolnikova M, Ding B, Rönnelid J, Alfredsson L, et al. Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheumatology 2009;60:924–30.
[13] Skinningsrud B, Lie BA, Husebye ES, Kvien TK, Førre Ø, Flatø B, et al. A CLEC16A variant confers risk for juvenile idiopathic arthritis and anti-cyclic citrullinated peptide antibody negative rheumatoid arthritis. Annals of the Rheumatic Diseases 2010;69:1471–4.
[14] Sigurdsson S, Padyukov L, Huizinga TW, Alm G, Syvanen AC, Ronnblom L. Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. Arthritis Rheumatology 2007; 56:2202–10.
[15] Garnier S, Dieude P, Michou L, Barbet S, Tan A, Lasbleiz S, et al. IRF5rs2004640-T allele, the new genetic factor for systemic lupus erythematosus, is not associated with rheumatoid arthritis. Annals of Rheumatic Diseases 2007;66: 828–31.
*[16] Padyukov L, Seielstad M, Ong RT, Ding B, Rönnelid J, Seddighzadeh M, et al. Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study group. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. Annals of Rheumatic Diseases 2011;70:259–65.

[17] van der Helm-van Mil AH, Kern M, Gregersen PK, Huizinga TW. Variation in radiologic joint destruction in rheumatoid arthritis differs between monozygotic and dizygotic twins and pairs of unrelated patients. Arthritis Rheum 2006;54:2028–30.
[18] Gonzalez-Gay MA, Garcia-Porrua C, Hajeer AH. Influence of human leukocyte antigen-DRB1 on the susceptibility and severity of rheumatoid arthritis. Seminars in Arthritis Rheumatology 2002;31:355–60.
[19] Wagner U, Kaltenhäuser S, Sauer H, Arnold S, Seidel W, Häntzschel H, et al. HLA markers and prediction of clinical course and outcome in rheumatoid arthritis. Arthritis Rheumatology 1997;40:341–51.
[20] Marinou I, Healy J, Mewar D, Moore DJ, Dickson MC, Binks MH, et al. Association of interleukin-6 and interleukin-10 genotypes with radiographic damage in rheumatoid arthritis is dependent on autoantibody status. Arthritis Rheumatology 2007;56:2549–56.
[21] Karlson EW, Chibnik LB, Cui J, Plenge RM, Glass RJ, Maher NE, et al. Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study. Annals of Rheumatic Diseases 2008;67:358–63.
[22] Cantagrel A, Navaux F, Loubet-Lescoulié P, Nourhashemi F, Enault G, Abbal M, et al. Interleukin-1beta, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis. Arthritis Rheumatology 1999;42:1093–100.
[23] Buchs N, di Giovine FS, Silvestri T, Vannier E, Duff GW, Miossec P. IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: interaction with their plasma levels. Genes and Immunology 2001;2:222–8.
[24] Plant D, Thomson W, Lunt M, Flynn E, Martin P, Eyre S, et al. The role of rheumatoid arthritis genetic susceptibility markers in the prediction of erosive disease in patients with early inflammatory polyarthritis: results from the Norfolk Arthritis Register. Rheumatology 2011;50:78–84.
[25] van der Linden MP, Feitsma AL, le Cessie S, Kern M, Olsson LM, Raychaudhuri S, et al. Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis. Arthritis Rheumatology 2009;60:2242–7.
*[26] Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheumatology 2004;50:3085–92.
[27] Hazes JM, Dijkmans BA, Vandenbroucke JP, de Vries RR, Cats A. Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. Annals of Rheumatic Diseases 1990;49:980–2.
[28] Voigt LF, Koepsell TD, Nelson JL, Dugowson CE, Daling JR. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. Epidemiology 1994;5:525–32.
*[29] Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Annals of Rheumatic Diseases 2010;69:70–81.
[30] Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. American Journal of Medicine 2006;119(503):e1–9.
[31] Georgiou TO, Marshall RI, Bartold PM. Prevalence of systemic diseases in Brisbane general and periodontal practice patients. Australian Dentel Journal 2004;49:177–84.
[32] de Pablo P, Dietrich T, McAlindon TE. Association of periodontal disease and tooth loss with rheumatoid arthritis in the US population. Journal of Rheumatology 2008;35:70–6.
[33] Kvien TK, Uhlig T, Ødegård S, Heiberg MS. Epidemiological aspects of rheumatoid arthritis: the sex ratio. Annals of the New York Academy of Sciences 2006;1069:212–22.
[34] Silman A, Kay A, Brennan P. Timing of pregnancy in relation to the onset of rheumatoid arthritis. Arthritis Rheumatology 1992;35:152–5.
[35] Kellgren JH, Lawrence JS, Aitken-Swan J. Rheumatic Complaints in and Urban population. Annals of Rheumatic Diseases 1953;12:5–15.
[36] Gabriel SE, Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955-1985. Arthritis Rheumatology 1999;42:415–20.
[37] Lindstrom TM, Robinson WH. Rheumatoid arthritis: a role for immunosenescence? Journal of American Geriatrics Society 2010;58:1565–75.
*[38] Källberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Annals of Rheumatic Diseases 2009;68:222–7.
[39] Berglin E, Kokkonen H, Einarsdottir E, Agren A, Rantapää Dahlqvist S. Influence of female hormonal factors, in relation to autoantibodies and genetic markers, on the development of rheumatoid arthritis in northern Sweden: a case-control study. Scandavian Journal of Rheumatology 2010;39:454–60.
[40] Pikwer M, Bergström U, Nilsson JA, Jacobsson L, Berglund G, Turesson C. Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis. Annals of Rheumatic Diseases 2009;68:526–30.
[41] Ginanjar E, Sumariyono, Setiati S, Setiyohadi B. Vitamin D and autoimmune disease. Acta Med Indones 2007;39:133–41.
[42] Merlino LA, Curtis J, Mikuls TR, Cerhan JR, Criswell LA, Saag KG, et al. Vitamin D intake is inversely associated with rheumatoid arthritis: results from the Iowa Women's Health Study. Arthritis Rheumatology 2004;50:72–7.
[43] Costenbader KH, Feskanich D, Holmes M, Karlson EW, Benito-Garcia E. Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women. Annals of Rheumatic Diseases 2008;67:530–5.
[44] Paglieroni TG, Ward J, Holland PV. Changes in peripheral blood CD5 (Bla) B-cell populations and autoantibodies following blood transfusion. Transfusion 1995;35:189–98.
[45] Symmons DP, Bankhead CR, Harrison BJ, Brennan P, Barrett EM, Scott DG, et al. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. Arthritis Rheumatology 1997;40:1955–61.
[46] Cerhan JR, Saag KG, Criswell LA, Merlino LA, Mikuls TR. Blood transfusion, alcohol use, and anthropometric risk factors for rheumatoid arthritis in older women. Journal of Rheumatology 2002;29:246–54.
[47] Bengtsson C, Nordmark B, Klareskog L, Lundberg I, Alfredsson L. EIRA Study Group. Socioeconomic status and the risk of developing rheumatoid arthritis: results from the Swedish EIRA study. Annals of Rheumatic Diseases 2005;64:1588–94.

[48] Pedersen M, Jacobsen S, Klarlund M, Frisch M. Socioeconomic status and risk of rheumatoid arthritis: a Danish case-control study. Journal of Rheumatology 2006;33:1069–74.
[49] Mandl LA, Costenbader KH, Simard JF, Karlson EW. Is birth weight associated with risk of rheumatoid arthritis? Data from a large cohort study. Annals of Rheumatic Diseases 2009;68:514–8.
[50] Karlson EW, Mandl LA, Hankinson SE, Grodstein F. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheumatology 2004;50:3458–67.
[51] Brennan P, Silman A. Breast-feeding and the onset of rheumatoid arthritis. Arthritis Rheumatology 1994;37:808–13.
[52] Brennan P, Ollier B, Worthington J, Hajeer A, Silman A. Are both genetic and reproductive associations with rheumatoid arthritis linked to prolactin? Lancet 1996;348:106–9.
[53] Jorgensen C, Picot MC, Bologna C, Sany J. Oral contraception, parity, breast feeding, and severity of rheumatoid arthritis. Annals of Rheumatic Diseases 1996;55:94–8.
[54] Assche EL, Susanne C. Increase in the amount of fetal lymphocytes in maternal blood during pregnancy. Journal of Medical Genetics 1980;17:267–72.
[55] Feitsma AL, van der Helm-van Mil AH, Huizinga TW, de Vries RR, Toes RE. Protection against rheumatoid arthritis by HLA: nature and nurture. Annals of Rheumatic Diseases 2008;67(Suppl 3):iii61–3.
[56] Firestein GS, Alvaro-Gracia JM, Maki R. Quantitative analysis of cytokine gene expression in rheumatoid arthritis. Journal of Immunology 1990;144:3347–53.
[57] Burmester GR, Stuhlmüller B, Keyszer G, Kinne RW. Mononuclear phagocytes and rheumatoid synovitis. Mastermind or workhorse in arthritis? Arthritis Rheumatology 1997;40:5–18.
[58] Genovese MC, Becker J, Schiff M, Luggen M, Sherrer Y, Kremer J, et al. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor. New England Journal of Medical 2005;353:1114–23.
[59] Andersson AK, Li C, Brennan FM. Recent developments in the immunobiology of rheumatoid arthritis. Arthritis Research and Therapy 2008;10:204.
[60] Sebbag M, Parry SL, Brennan FM, Feldmann M. Cytokine stimulation of T lymphocytes regulates their capacity to induce monocyte production of tumor necrosis factor-alpha, but not interleukin-10: possible relevance to pathophysiology of rheumatoid arthritis. European Journal of Immunology 1997;27:624–32.
[61] Kotake S, Udagawa N, Takahashi N, Matsuzaki K, Itoh K, Ishiyama S, et al. IL-17 in synovial fluids from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. Journal of Clinical Investication 1999;103:1345–52.
[62] Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, et al. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. Arthritis Rheumatology 2006;54:2793–806.
[63] Martinez-Gamboa L, Brezinschek HP, Burmester GR, Dorner T. Immunopathologic role of B lymphocytes in rheumatoid arthritis: rationale of B cell-directed therapy. Autoimmunity Reviews 2006;5:437–42.
[64] Carroll MC. The complement system in regulation of adaptive immunity. Nature Immunology 2004;5:981–6.
[65] Schroder AE, Greiner A, Seyfert C, Berek C. Differentiation of B cells in the nonlymphoid tissue of the synovial membrane of patients with rheumatoid arthritis. Proceedings of National of Academy of Sciences U S A 1996;93:221–5.
[66] Bartok B, Firestein GS. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. Immunology Reviews 2010; 233:233–55.
[67] Wegner N, Lundberg K, Kinloch A, Fisher B, Malmström V, Feldmann M, et al. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. Immunology Reviews 2010;233:34–54.
[68] Senshu T, Akiyama K, Kan S, Asaga H, Ishigami A, Manabe M. Detection of deiminated proteins in rat skin: probing with a monospecific antibody after modification of citrulline residues. Journal of Investigative Dermatology 1995;105:163–9.
[69] Moscarello MA, Wood DD, Ackerley C, Boulias C. Myelin in multiple sclerosis is developmentally immature. Journal of Clinical Investigation 1994;94:146–54.
[70] Makrygiannakis D, af Klint E, Lundberg IE, Löfberg R, Ulfgren AK, Klareskog L, et al. Citrullination is an inflammation-dependent process. Annals of Rheumatic Diseases 2006;65:1219–22.
[71] Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, et al. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. Annals of Internal Medicine 2007;146:797–808.
*[72] Nielen MMJ, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MHMT, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheumatology 2004;50:380–6.
[73] Rantapää-Dahkqvist S, de Jong BAW, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor Predict the development of rheumatoid arthritis. Arthritis Rheumatology 2003;48:2741–9.
[74] Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, et al. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Annals of Rheumatic Diseases 2008;67:1488–92.
[75] Quirke AM, Fisher BA, Kinloch AJ, Venables PJ. Citrullination of autoantigens: Upstream of TNFα in the pathogenesis of rheumatoid arthritis. FEBS Letters; 2011 [Epub ahead of print].
[76] Mikuls TR, Payne JB, Reinhardt RA, Thiele GM, Maziarz E, Cannella AC, et al. Antibody responses to Porphyromonas gingivalis (P. gingivalis) in subjects with rheumatoid arthritis and periodontitis. International Immunopharmacology 2009;9:38–42.
[77] Feitsma AL, van der Voort EI, Franken KL, el Bannoudi H, Elferink BG, Drijfhout JW, et al. Identification of citrullinated vimentin peptides as T cell epitopes in HLA-DR4-positive patients with rheumatoid arthritis. Arthritis Rheumatology 2010;62:117–25.
[78] Hill JA, Bell DA, Brintnell W, Yue D, Wehrli B, Jevnikar AM, et al. Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in DR4-IE transgenic mice. Journal of Experimental Medicine 2008;205:967–79.
[79] van der Woude D, Rantapää-Dahlqvist S, Ioan-Facsinay A, Onnekink C, Schwarte CM, Verpoort KN, et al. Epitope spreading of the anti-citrullinated protein antibody response occurs before disease onset and is associated with the disease course of early arthritis. Annals of Rheumatic Diseases 2010;69:1554–61.

[80] Klareskog L, Padyukov L, Rönnelid J, Alfredsson L. Genes, environment and immunity in the development of rheumatoid arthritis. Current Opinion Immunology 2006;18:650–5.
[81] Clavel C, Nogueira L, Laurent L, Iobagiu C, Vincent C, Sebbag M, et al. Induction of macrophage secretion of tumor necrosis factor alpha through Fcgamma receptor IIa engagement by rheumatoid arthritis specific autoantibodies to citrullinated proteins complexed with fibrinogen. Arthritis Rheumatology 2008;58:678–88.
[82] Pedersen M, Jacobsen S, Garred P, Madsen HO, Klarlund M, Svejgaard A, et al. Strong combined gene-environment effects in anti-cyclic citrullinated peptide-positive rheumatoid arthritis: a nationwide case-control study in Denmark. Arthritis Rheumatology 2007;56:1446–53.
[83] Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH, et al. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. American Journal of Human Genetics 2007;80:867–75.
*[84] Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, Ding B, et al. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Annals of Rheumatic Diseases 2010;69:1077–85.
[85] Ma MH, Scott IC, Kingsley GH, Scott DL. Remission in early rheumatoid arthritis. Journal of Rheumatology 2010;37:1444–53.
[86] Bykerk VP. Strategies to prevent rheumatoid arthritis in high-risk patients. Current Opinion Rheumatology 2011;23:179–84.
[87] Bos WH, Dijkmans BA, Boers M, van de Stadt RJ, van Schaardenburg D. Effect of dexamethasone on autoantibody levels and arthritis development in arthralgia patients: a randomized trial. Annals of Rheumatic Diseases 2010;69:571–4.
[88] Verstappen SM, McCoy MJ, Roberts C, Dale NE, Hassell AB, Symmons DP. STIVEA investigators. The beneficial effects of a 3-week course of intramuscular glucocorticoid injections in patients with very early inflammatory polyarthritis: results of the STIVEA trial. Annals of Rheumatic Diseases 2010;69:503–9.
[89] van Dongen H, van Aken J, Lard LR, Visser K, Ronday HK, Hulsmans HM, et al. Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. Arthritis Rheumatology 2007;56:1424–32.
[90] Emery P, Durez P, Dougados M, Legerton CW, Becker JC, Vratsanos G, et al. Impact of T-cell costimulation modulation in patients with undifferentiated inflammatory arthritis or very early rheumatoid arthritis: a clinical and imaging study of abatacept (the ADJUST trial). Annals of Rheumatic Diseases 2010;69:510–6.
[91] Saleem B, Mackie S, Quinn M, Nizam S, Hensor E, Jarrett S, et al. Does the use of tumour necrosis factor antagonist therapy in poor prognosis, undifferentiated arthritis prevent progression to rheumatoid arthritis? Annals of Rheumatic Diseases 2008;67:1178–80.
[92] de Vries RR, van der Woude D, Houwing JJ, Toes RE. Genetics of ACPA-positive rheumatoid arthritis: the beginning of the end? Annals of Rheumatic Diseases 2011;70(Suppl 1):i51–4.
[93] Lee HS, Korman BD, Le JM, Kastner DL, Remmers EF, Gregersen PK, et al. Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. Arthritis & Rheumatism 2009;60:364–71.
[94] Hughes LB, Reynolds RJ, Brown EE, Kelley JM, Thomson B, Conn DL, et al. Most common single-nucleotide polymorphisms associated with rheumatoid arthritis in Persons of European ancestry confer risk of rheumatoid arthritis in African Americans. Arthritis Rheumatology 2010;62:3547–53.
[95] Visser K, Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Ronday HK, Seys PE, Kerstens PJ, et al. A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis receiving different dynamic treatment strategies: post hoc analyses from the BeSt study. Annals of Rheumatic Diseases 2010;69:1333–7.